

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO**  
**CENTRO DE CIENCIAS DA SAÚDE**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**

**LEIDY JOHANA MADROÑERO**

**ANÁLISE TRANSCRIPTÔMICA DA INTERAÇÃO**  
**MAMOEIRO-*Papaya Meleira Virus***

VITÓRIA-ES

2014

LEIDY JOHANA MADROÑERO

**ANÁLISE TRANSCRIPTÔMICA DA INTERAÇÃO  
MAMOEIRO-*Papaya Meleira Virus***

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Biotecnologia.

Orientadora: Prof. Dra. Patricia Machado Bueno Fernandes.

Co-orientador: Prof. Dr. Antonio Alberto Ribeiro Fernandes

VITÓRIA-ES

2014

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

---

Madroñero, Leidy Johana, 1988-  
M178a      Análise transcritômica da interação mamoeiro-Papaya  
              Meleira Virus / Leidy Johana Madroñero – 2014.  
              76 f. : il.

Orientador: Patricia Machado Bueno Fernandes.  
Coorientador: Antonio Alberto Ribeiro Fernandes.

Dissertação (Mestrado em Biotecnologia) – Universidade  
Federal do Espírito Santo, Centro de Ciências da Saúde.

1. RNA-seq. 2. Genes. 3. Mamoeiro. I. Fernandes, Patricia  
Machado Bueno. II. Fernandes, Antonio Alberto Ribeiro.  
III. Universidade Federal do Espírito Santo. Centro de Ciências  
da Saúde. IV. Título.

CDU: 61

---

**LEIDY JOHANA MADROÑERO**

**ANÁLISE TRANSCRIPTÔMICA DA INTERAÇÃO MAMOEIRO-*Papaya Meleira*  
*Virus***

Dissertação apresentada ao Programa de Pós Graduação em Biotecnologia do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do título de Mestre em Biotecnologia.

Aprovada em 27 de Novembro de 2014.

---

**Prof. Dra. Patricia Machado Bueno Fernandes**  
**Universidade Federal do Espírito Santo**  
**Orientadora**

---

**Prof. Dr. Antonio Alberto Ribeiro Fernandes**  
**Universidade Federal do Espírito Santo**  
**Coorientador**

---

**Prof. Dr. José Aires Ventura**  
**Universidade Federal do Espírito Santo**  
**Membro Interno**

---

**Prof. Dra. Daisy de La Caridad Pérez Brito**  
**CICY**  
**Membro Externo**

---

**Prof. Dr. Raúl Tapia Tussell**  
**CICY**  
**Membro Externo**

Vitória  
2014

## DEDICATÓRIA

*A minha mãe, a minha irmã e ao meu irmão, por serem "mi polo a tierra" e minha principal motivação para continuar nos momentos mais críticos.*

*A Alexander, porque mesmo como amigo ou como companheiro, de perto ou de longe, tem sido sempre meu "suporte" e minha inspiração para ser uma pessoa melhor.*

*Aos meus amigos, que se tornam minha família, enquanto estou longe de casa.*

## **AGRADECIMENTOS**

Ao programa de bolsistas da OEA-GCUB, 2012 por ter-me concedido a oportunidade de cursar o mestrado em Biotecnologia na Universidade Federal do Espírito Santo.

À agência de apoio CAPES pelo financiamento de minha bolsa de estudos.

Às agências de apoio e financiamento: CNPq, FAPES, CAPES e FINEP pelo financiamento de reagentes, materiais e serviços requeridos para o desenvolvimento do projeto.

À Prof. Dra. Patricia M. B. Fernandes pela dedicação, orientação e pelo exemplo de determinação.

Ao professor Alberto e ao professor Aires pelo suas sugestões, questionamentos e correções, que contribuíram para melhorar a qualidade do trabalho.

Ao Eduardo, Silas, Tathiana, Paolla e Oeber, que contribuíram com seus conhecimentos e suas experiências ao longo de meu trabalho.

Agradeço às funcionárias da secretaria do programa de pós-graduação, especialmente a Kárita pela atenção e a gentileza.

A todos os colegas de trabalho do Laboratório de Biotecnologia Aplicada ao Agronegócio que me acolheram e copartilharam comigo sua linguagem, sua experiência de vida, sua trajetória e demais detalhes que me fizeram crescer e aprender sobre vocês e sobre sua cultura.

"I believe that we have been doing this not primarily to achieve riches or even honour, but rather because we were interested in the work, enjoyed doing it and felt very strongly that it was worthwhile."

Frederick Sanger

## RESUMO

O mamoeiro (*Carica papaya* L.) é uma das fruteiras mais cultivadas nas regiões tropicais e subtropicais do mundo. O Brasil faz parte do grupo dos países que mais produzem e exportam mamão no mundo. O Espírito Santo e a Bahia são responsáveis por mais de 70% da área brasileira produtora deste fruto. Porém, doenças causadas por microrganismos infecciosos afetam de modo considerável sua produção. Entre as principais doenças, destaca-se a meleira do mamoeiro, causada pelo *Papaya meleira virus* (PMeV), que ainda não possui uma cultivar resistente. Interessantemente os sintomas somente são desencadeados após a frutificação. Os mecanismos moleculares envolvidos no desenvolvimento dos sintomas e na resposta de defesa da planta ao PMeV ainda não foram esclarecidos. Para entender os pontos chaves desta interação, que permitam o desenvolvimento de metodologias de melhoramento genético, um estudo transcriptômico foi abordado. A tecnologia *RNA-seq* foi usada para o sequenciamento do transcriptoma a partir de plantas com 3, 6 e 8 meses de idade após plantio, inoculadas e não inoculadas com o PMeV. Os genes diferencialmente expressos nos 3 tempos e nas duas condições foram preditos e analisados. Estas análises revelaram um padrão de expressão geral dos genes envolvidos nesta interação. Foram encontrados 21 genes com o perfil de expressão alterado nas plantas inoculadas exclusivamente nos seis meses de idade. Destes, 8 genes envolvidos em processos de respostas de defesa e morte celular, resposta ao estresse e resposta ao estímulo biótico e abiótico foram reprimidos; enquanto os demais (13 genes), envolvidos principalmente em processos metabólicos primários, biogêneses, diferenciação e ciclo celular, comunicação e crescimento celular, bem como processos envolvidos em reprodução, e desenvolvimento da floração, foram superexpressos. Estes resultados sugerem que, aos seis meses de idade, a planta é obrigada a alterar seu programa de expressão gênica, direcionando a resposta para os processos próprios do desenvolvimento, requeridos nesse estágio fisiológico, que primam sob a resposta ao estresse, fato que finalmente leva ao desenvolvimento dos sintomas.

Palavras chave: Interação mamão-PMeV, *RNA-seq*, expressão diferencial de genes.



## ABSTRACT

Papaya is one of the fruit crops most cultivated in tropical and subtropical regions. Brazil is a major producer and exporter of papaya in the world. The largest area in Brazil, about 70%, for producing papaya is located in Espiritu Santo and Bahia. However this production is affected by infectious diseases caused by pathogens. The sticky disease caused by *Papaya meleira virus* (PMeV) is one of the most severe diseases. No resistance has been reported for sticky disease and interestingly their symptoms only are triggered at the ripening. The molecular mechanisms involved in both the symptoms' development and in the papaya defense response are still unclear. To understand the key point in this pathosystem leading to papaya crop genetic improvement methodologies we conducted a transcriptomics study. RNA-seq technology was used to sequencing the transcriptome from PMeV inoculated and non inoculated plants with 3, 6 and 8 months old. The differentially expressed genes in the both conditions and in the three times were found. Using different graphics analysis we show the global gene expression patterns in this interaction. We found 21 genes exhibit an altered profile at six month just in the inoculated condition. 8 genes related with defense response like cellular death and stress responses and biotic and abiotic stimulus were down regulated whereas 13 genes involved with primary metabolic process, biogenesis, cell differentiation, cell cycle, cell communication, cell growth, as well as in reproduction and flower development were up regulated. This results suggest that in the six month the plant is forced to change their gene expression program routed to response for the physiological processes involved just at this period and should this is being favored over the stress response leading to the symptoms development.

Key words: *C. papaya*-PMeV interaction, RNA-seq, differentially expressed genes.

## LISTA DE FIGURAS

Figura 1: Respostas de defesa envolvidas na interação planta-vírus análogas às respostas de defesa contra bactérias e fungos. ....	21
Figura 2: Metodologia geral usada na técnica de sequenciamento RNA-seq. ....	25
Figura 3: Fluxograma seguido no desenvolvimento da metodologia agrupado em três etapas principais mostradas em cor vermelho. ....	35
Figura 4: Plantas Inoculadas com látex infectado com PMeV diluído em tampão fosfato de sódio 50Mm pH 7,0 (1:1).. ....	36
Figura 5: Plantas inoculadas com tampão fosfato de sódio 50 Mm pH 7.0. ....	37
Figura 6: (a) Ligaç�o das sequ�ncias adaptadores aos fragmentos de DNA obtidos a partir da constru��o da biblioteca de cDNA e posteriores modifica��es.. ....	40
Figura 7: Amplifica��o em ponte.. ....	42
Figura 8: Diagn�stico de meleira.....	45
Figura 9: Gr�fico de escalonamento multidimensional (MDS).....	51
Figura 10: <i>Heat map</i> representando as dist�ncias <i>Jensen-Shannon</i> (JS) dos n�veis de express�o entre as todas amostras sequenciadas.....	52
Figura 11: Gr�fico de densidade. Densidade de genes em fun��o do seus n�veis de express�o dados pela raz�o Log2 (Fold Change) para todas as condi��es. ....	55
Figura 12: Diagrama de caixas.....	56
Figura 13: Diagramas de Venn mostrando as rela��es entre os diferentes conjuntos de genes diferencialmente expressos. ....	60
Figura 14: Diagrama de cluster. ....	61
Figura 15: Anota��o funcional GO dos genes.....	62

## LISTA DE TABELAS

Tabela 1. Resumo comparativo entre as principais plataformas de sequenciamento. Fonte: Adaptada de: Barba, 2014.....	31
Tabela 2: Observações feitas das plantas em cada ponto avaliado. ....	34
Tabela 3: Delineamento experimental .....	34
Tabela 4: Controle de qualidade do RNA. ....	39
Tabela 5: Diagnostico da presença de PMeV em plantas de <i>C. papaya</i> . ....	45
Tabela 6: Estatística das leituras geradas e seu mapeamento no genoma de <i>C. papaya</i> .....	47
Tabela 7: Genes diferencialmente expressos (DEGs) em plantas não Inoculadas com PMeV aos 3,6 e 8 meses de idade quando comparadas com plantas de 2 meses de idade, usando um p-value ajustado a uma FDR=0,05 (q-value). ....	53
Tabela 8: Genes diferencialmente expressos (DEGs) em plantas Inoculadas com PMeV aos 3,6 e 8 meses de idade quando comparadas com plantas de 2 meses de idade, usando um p-value ajustado a uma FDR=0,05 (q-value). ....	53
Tabela 9: Sumario da distribuição estadística dos dados no diagrama de caixa. ....	56

## LISTA DE SIGLAS

cDNA	DNA complementar (do inglês <i>Complementary DNA</i> )
dsRNA	RNA dupla-fita (do inglês <i>Double-strand RNA</i> )
FDR	Taxa de falsa descoberta (do inglês <i>False discovery rate</i> )
FPKM	Fragmentos por kilobase de transcrito por milhões de fragmentos mapeados (do inglês <i>Fragments Per Kilobase of transcript per Million fragments mapped</i> )
GO	Ontologia Gênica (do inglês <i>Gene ontology</i> )
HR	Reação de hipersensibilidade (do inglês <i>Hypersensitive response</i> )
JA	Ácido jasmônico (do inglês <i>Jasmonic acid</i> )
miRNAs	microRNAs
NGS	Sequenciamento de nova geração (do inglês <i>Next Generation Sequencing</i> )
NO	Óxido nítrico (do inglês <i>Nitric oxide</i> )
PMeV	<i>Papaya meleira virus</i>
RNA seq	Sequenciamento do RNA (do inglês <i>RNA sequencing</i> )
ROS	Espécies reativas de oxigênio (do inglês <i>Reactive oxygen species</i> )

RT-PCR	Transcrição reversa-Reação em cadeia da polimerase (do inglês <i>Reverse transcription-Polymerase chain reaction</i> )
SA	Ácido salicílico (do inglês <i>Salicylic acid</i> )
SAR	Resposta sistêmica adquirida (do inglês <i>Systemic acquired resistance</i> )
UPS	Sistema Ubiquitina proteasome (do inglês <i>Ubiquitin proteasome system</i> )

## SUMÁRIO

1. INTRODUÇÃO .....	14
1.1. Importância econômica da cultura do mamoeiro.....	14
1.2. A meleira do mamoeiro .....	14
1.3 Interações planta-patógeno: <i>C.papaya</i> -PMeV .....	17
1.4 Respostas de defesa envolvidas na infecção viral .....	18
1.4.1. Resposta de Hipersensibilidade e Necroses .....	18
1.4.2 Resposta de Necroses sistêmica.....	19
1.4.3. Resposta sistêmica adquirida (SAR) .....	20
1.4.4. Sistema Ubiquitina-proteosoma (UPS) .....	22
1.5. RNA-seq no estudo de interações planta-patógeno.....	23
1.6. Bioinformática no processamento de dados gerados a partir da <i>Illumina</i> .....	27
2. OBJETIVOS: .....	32
2.1 Objetivo geral: .....	32
2.2 Objetivos específicos: .....	32
3. MATERIAIS E MÉTODOS.....	33
3.2. Inoculação das plantas de mamoeiro.....	33
3.3. Coleta e armazenamento .....	33
3. 4. Delineamento experimental.....	34
3.4. Extração de RNA e diagnóstico molecular do PMeV: .....	38

3.5. Testes de qualidade do RNA: .....	38
3.6. Construção da Biblioteca de cDNA. ....	39
3.8. Sequenciamento .....	41
3.9. Análises de bioinformática .....	43
3.9.1. Verificação da qualidade das leituras geradas .....	43
3.9.2 Alinhamento das leituras, montagem de transcritos e análises de expressão diferencial de genes. ....	43
3.10. Análises de dados e gráficos de expressão .....	44
4. RESULTADOS E DISCUSSÃO .....	45
4.1. Diagnóstico do PMeV nas amostras selecionadas para o sequenciamento ...	45
4.2. Sequenciamento .....	46
4.3. Correlação entre as amostras .....	48
4.4. Análises de expressão diferencial de genes .....	49
4.5. Gráficos de Análises de expressão diferencial de genes .....	54
5. CONCLUSÃO .....	65
6. REFERÊNCIAS .....	66

## **1. INTRODUÇÃO**

### **1.1. Importância econômica da cultura do mamoeiro**

O mamoeiro é uma das principais culturas das regiões tropical e subtropical e seu fruto é conhecido por seu alto valor nutricional e suas diversas aplicações medicinais.

A produção mundial de mamão é superior a 12.4 milhões de toneladas, sendo a Índia, o Brasil, a República Dominicana, a Indonésia e a Nigéria os principais países produtores. O México, a Malásia e o Brasil são os maiores exportadores (FAO, 2014).

O Brasil encerrou o ano de 2012 com uma produção de 1,52 milhões de toneladas (t) (FAO, 2014) sendo considerado o segundo maior produtor de mamão. As exportações brasileiras de mamão geraram no ano de 2011 um rendimento de US\$ 38,8 milhões (FAO, 2013).

O Espírito Santo é o segundo estado brasileiro com maior produção de mamão e, junto com a Bahia responde por 70% da área e da produção do país. A cultura do mamoeiro encontra-se entre as principais atividades agrícolas do estado e a cadeia de produção e comercialização gera uma renda bruta anual da ordem de R\$ 75 milhões, empregando cerca de 30 mil pessoas (MARTINS *et al.*, 2008). A produção concentra-se em 11 municípios no norte do estado, sendo Linhares a mais importante exportadora de mamão do Brasil.

A produtividade de mamão, no entanto, é afetada principalmente por doenças infecciosas, as quais são responsáveis por perdas econômicas significativas na produção, venda e exportação da fruta fresca. Dentre estas destaca-se a meleira do mamoeiro.

### **1.2. A meleira do mamoeiro**

A meleira do mamoeiro foi relatada na década de 1980 nos estados brasileiros da



Bahia e do Espírito Santo. Em poucos anos a doença expandiu-se aos estados de Pernambuco e do Ceará. Atualmente esta doença ocorre no Espírito Santo, Bahia, Pernambuco, Ceará, Rio Grande do Norte e Minas Gerais (DALTRO *et al.*, 2014).

Primeiramente, os sintomas da meleira foram atribuídos a fatores abióticos, como a um distúrbio na absorção de Cálcio e Boro, associado a estresse hídrico resultante da falta de água no solo (NAKAGAWA *et al.*, 1987) e ao desbalanceamento de bases trocáveis no solo (CORREA *et al.*, 1988).

A natureza biótica da meleira foi confirmada quando plantas inicialmente sadias apresentaram os sintomas típicos da doença após serem inoculados com látex coletado a partir de plantas doentes, sugerindo que o agente causador da doença estivesse presente no látex das plantas (RODRIGUES *et al.*, 1989; KITAJIMA *et al.*, 1993).

A etiologia viral da meleira foi confirmada após a purificação das partículas virais presente no látex, seguida de inoculação em mamoeiros sadios, que se tornaram doentes. Assim, o vírus da meleira do mamoeiro, *Papaya meleira virus* (PMeV), é um vírus de partícula isométrica com aproximadamente 50 nm de diâmetro e genoma composto por uma única molécula de RNA dupla-fita (dsRNA) com aproximadamente 12.000 (pb) em gel de agarose (MACIEL-ZAMBOLIM *et al.*, 2003).

Na ocorrência da meleira, verifica-se a exsudação espontânea de látex aquoso e fluido de frutos e folhas, consequência da ruptura dos laticíferos pelo aumento da turgência e alterações morfológicas do látex derivadas da presença do vírus (VENTURA *et al.*, 2003; RODRIGUES *et al.*, 2009). Em contato com o ar, o látex é oxidado e causa pequenas lesões necróticas nas pontas de folhas jovens e à coloração escura e “melada” no fruto (VENTURA *et al.*, 2004). Dados de microscopia eletrônica e dados moleculares indicam que as partículas virais estão fortemente ligadas aos polímeros presentes no látex, possivelmente como um mecanismo de proteção ou para auxiliar o transporte viral (RODRIGUES *et al.*, 2005; RODRIGUES *et al.*, 2009). Em estádios mais avançados da doença, também é possível visualizar a presença de manchas zonadas superficiais de cor verde-clara (VENTURA *et al.*, 2001).

Na presença da doença, o sabor e a consistência dos frutos é alterado de tal maneira que não se cumprem os requisitos de qualidade para o comércio (VENTURA e COSTA, 2003). Além disso, a meleira aumenta a suscetibilidade dos frutos à infestação pelas moscas-das-frutas, *Ceratitis capitata*, que, quando encontrada numa concentração acima da permitida pelas normas norte-americanas, (APHIS/USDA) inviabilizam a exportação de mamão (MARTINS, 2003; MARTINS *et al.*, 2012).

A identificação dos sintomas da doença e a subsequente erradicação das plantas doentes (roguing) é atualmente a melhor estratégia para controlar a meleira do mamoeiro. Geralmente, o vírus infecta pelo menos 20% das plantas durante o ciclo produtivo, em algumas plantações, nas quais o roguing não é feito, têm-se registrado uma incidência de até 100% na fase de colheita em plantas com apenas 12-15 meses (VENTURA *et al.*, 2004).

Entretanto, os sintomas são disparados somente após a floração e, portanto, plantas infectadas livres dos sintomas são capazes de transmitir o PMeV (VENTURA *et al.*, 2004). Diante disso, surgiu a necessidade do desenvolvimento de um método de diagnose precoce.

O primeiro método molecular que permitiu a detecção do PMeV foi realizado através da extração do dsRNA viral a partir do látex das plantas sintomáticas para a meleira (RODRIGUES *et al.*, 2005). Entretanto, este método de detecção exige que a planta infectada tenha uma alta carga viral para a visualização do dsRNA viral no gel de agarose. Assim, mais recentemente, foram propostos dois novos métodos de diagnóstico utilizando-se RT-PCR e PCR quantitativo em tempo real, que são mais sensíveis e por isso permitem o diagnóstico do PMeV antes do aparecimento dos sintomas, inclusive em diferentes tecidos do mamoeiro (ABREU *et al.*, 2012). Esses métodos de diagnóstico permitiram identificar diversas plantas infectadas pelo PMeV assintomáticas para a meleira no campo, mesmo quando estas plantas já passaram pelo processo de florescimento.

### 1.3 Interações planta-patógeno: *C.papaya*-PMeV

A interação planta-patógeno neste modelo tem sido o principal foco de estudo do grupo do Laboratório de Biotecnologia Aplicada ao Agronegócio da Universidade Federal do Espírito Santo, Brasil. Desta forma, aportes significativos que visam às possíveis respostas envolvidas neste patossistema têm sido propostas.

O acúmulo de cristais de oxalato de cálcio no látex e o aumento da produção de peróxido de hidrogênio ( $H_2O_2$ ) nos laticíferos e nas células companheiras das plantas infectadas com PMeV foi demonstrado. Resultados que sugerem a possível ativação de uma resposta de defesa nos laticíferos contra a infecção por PMeV que envolve vias de sinalização que são desencadeadas ou que culminam no aumento nos níveis de  $H_2O_2$  e de oxalato de cálcio (RODRIGUES *et al.*, 2009) foram estudadas por proteômica e por análises de microRNAs (RODRIGUES *et al.*, 2009).

Nesta interação foi também, avaliado o papel do óxido nítrico (NO, do inglês *nitric oxide*), um importante sinalizador envolvido em respostas ao estresse em plantas. O resultado mostrou que o NO induz o aumento de peroxidases, compostos fenólicos e carboidratos em plantas infectadas. Ativação destas respostas antioxidantes é importante na defesa da planta contra vírus (BUSS *et al.*, 2011).

Com o objetivo de compreender as vias regulatórias moduladas pelo PMeV, amostras de plantas doentes coletadas no campo foram analisadas utilizando abordagens complementares de proteômica usando eletroforese bi-dimensional, 2-DE e marcação fluorescente diferencial, DIGE, seguidas de espectrometria de massas, LC-MS/MS, revelaram várias proteínas expressas diferencialmente. Enquanto que proteínas que estão relacionadas ao metabolismo são reprimidas, a calreticulina e proteínas da via do proteassoma, que estão relacionadas ao estresse, são induzidas em resposta à infecção com PMeV (RODRIGUES *et al.*, 2009; RODRIGUES *et al.*, 2011).

Para observar os efeitos causados por este vírus nos laticíferos de plantas infectadas, outra abordagem usando proteômica comparativa de látex de mamoeiros sadios e infectados mostraram que o PMeV reduz os níveis de cisteíno-proteases (de tipo quimopapaína) (RODRIGUES *et al.*, 2012).

Recentemente, um estudo de miRNAs envolvidos no sistema UPS e nas vias de resposta a estresse foram avaliados durante a infecção pelo PMeV. Os resultados mostraram uma alteração na acumulação de miRNAs envolvidos no sistema UPS em plantas infectadas pelo PMeV, bem como a alteração em miRNAs que modulam genes envolvidos em respostas de defesa ao estresse, como a via de espécies reativas de oxigênio (ROS) (ABREU *et al.*, 2014).

#### **1.4 Respostas de defesa envolvidas na infecção viral**

Os vírus induzem uma variedade de respostas nas células hospedeiras (Figura 1), as quais são mediadas pela perturbação de diferentes vias de sinalização. Nas últimas décadas, o entendimento das funções das proteínas virais e da biologia das plantas, bem como o desenvolvimento das novas tecnologias para avaliar expressão gênica, têm permitido elucidar os possíveis eventos que ocorrem dentro de células hospedeiras susceptíveis e resistentes à infecção viral.

Por muitos anos se tem estudado o silenciamento de RNA como o principal sistema de defesa das plantas contra os vírus, no entanto as plantas também induzam outro tipo de respostas como a reação de hipersensibilidade (HR) e a resistência sistêmica adquirida (SAR). Todas estas respostas juntas limitam a infecção e confirmam resistência aos tecidos não infetados.

##### **1.4.1. Resposta de Hipersensibilidade e Necroses**

HR e necroses são respostas contra um amplo espectro de patógenos de plantas que incluem fungos, bactérias e vírus. Salvo alguns pontos, os mecanismos usados são similares. Durante a infecção viral, de um modo similar ao acontecido nas infecções não virais, HR é começada pela interação de moléculas virais com as proteínas R da célula hospedeira (Avr/R) o que induz alterações metabólicas nos níveis de hormônios relacionados com a defesa da planta tal como o ácido salicílico (AS), ácido jasmônico (JA) e óxido nítrico (NO), assim como a acumulação de espécies reativas de oxigênio (ROS) como  $O^{2-}$  e peróxido de hidrogênio, respostas

que são observadas tanto em tecidos infetados como em tecidos não infetados (CULVER e PADMANABHAN, 2007; CARR *et al.*, 2010; PALLAS e GARCIA, 2011; MANDADI e SCHOLTHOF, 2012).

No nível celular HR afeta a homeostase do cálcio  $\text{Ca}^{2+}$  e altera o potencial e permeabilidade da membrana. Mesmo que durante HR diversas proteinases da família das caspases são ativadas, entre elas enzimas que tem a ver com processamento vacuolar, estas enzimas agem como efetores de morte celular ou necroses durante a HR. Já no nível molecular e bioquímico, diversas vias de sinalização genética são ativadas, para induzir a expressão de múltiplas proteínas, incluindo as *mitogen-activated protein* (MING *et al.*) proteínas quinases, a jusante diversas proteínas relacionadas com a defesa da planta como as glucanases, quitinases e defensinas relacionadas com proteínas envolvidas em patogêneses são induzidas (MUR *et al.*, 2008).

#### **1.4.2 Resposta de Necroses sistêmica**

A necroses sistêmica é parecida à necroses comumente observada, que também é o resultado de uma morte celular não controlada ou não constitutiva (LORRAIN, 2003; MOEDER e YOSHIOKA, 2008). Porém, ao contrário da necrose associada à HR, a necroses sistêmica é manifestada muito mais tarde na infecção e é observada primariamente nos tecidos superiores não inoculados. Além disso, acredita-se que a necrose sistêmica não impossibilita a multiplicação do vírus ou movimentação sistêmica, levando a uma infecção susceptível (MANDADI e SCHOLTHOF, 2013).

Embora ainda não estejam muito bem caracterizados os mecanismos moleculares que fundamentam a necrose sistêmica, estudos recentes têm mostrado que apesar dos diferentes papéis entre a necrose sistêmica e a necrose associada à HR, estas duas respostas compartilham notáveis similitudes ao nível bioquímico e molecular. Para citar alguns, os dois tipos de necroses envolvem morte celular, alteram a expressão de genes relacionados com defesa da planta e induzem o acúmulo de ROS (XU e ROOSSINCK, 2000; KIM *et al.*, 2008; KOMATSU *et al.*, 2010; XU *et al.*, 2012).

(KOMATSU *et al.*, 2010) determinou genes que estão envolvidos em HR nas interações não compatíveis que também estão envolvidos na necrose sistêmica das infecções compatíveis, por exemplo, as respostas da necrose sistêmica dependem do complexo funcional SGT1/RAR1 e requer também a sinalização MAPKKK $\alpha$ /MEK2. Neste mesmo trabalho, o autor demonstra que o silenciamento de SGT1 e RAR1 promove a acumulação do vírus em plantas de *N. benthamiana* infectadas com o *potexvirus* PLAMV. Este resultado contrasta com o postulado de que a necrose sistêmica não impede o acúmulo viral e sugere que a necrose sistêmica pode promover a imunidade antiviral durante as interações planta-vírus de tipo compatível.

A análise de expressão global de genes quando é disparada a resposta de necrose sistêmica mostrou que as os padrões de expressão apresentados em HR e a necrose sistêmica possuem fortes similaridades (PACHECO *et al.*, 2012).

Diversos trabalhos apoiam a discussão de que a necrose sistêmica envolve mecanismos fisiológicos, moleculares e bioquímicos similares a HR, porém a importância biológica nas interações compatíveis ainda é ambígua (MANDADI e SCHOLTHOF, 2013).

#### **1.4.3. Resposta sistêmica adquirida (SAR)**

De modo similar ao que ocorre no HR, SAR é disparada em interações não compatíveis e envolve Avr e as proteínas R nas células infectadas primárias. Porém a resistência é transduzida aos tecidos não infectados que estão distantes. Embora os mecanismos exatos ainda não estejam definidos, esta resposta é iniciada quando um fator de virulência Avr é reconhecido por uma proteína R da planta o que resulta na acumulação de fitormônios como ácido salicílico (SA) e ácido jasmônico (JA) nos tecidos distantes (VLOT *et al.*, 2008).

Ao contrario que HR, SAR é uma resposta imune duradoura que aponta a proporcionar resistência para subseqüentes infecções nos tecidos distantes. Não é claro como SAR pode agir durante um longo tempo, no entanto, modificações epigenéticas, como a metilação do DNA ou a remodelagem da cromatina, podem ser

críticas para manter a sinal de SAR (SPOEL e DONG, 2012).

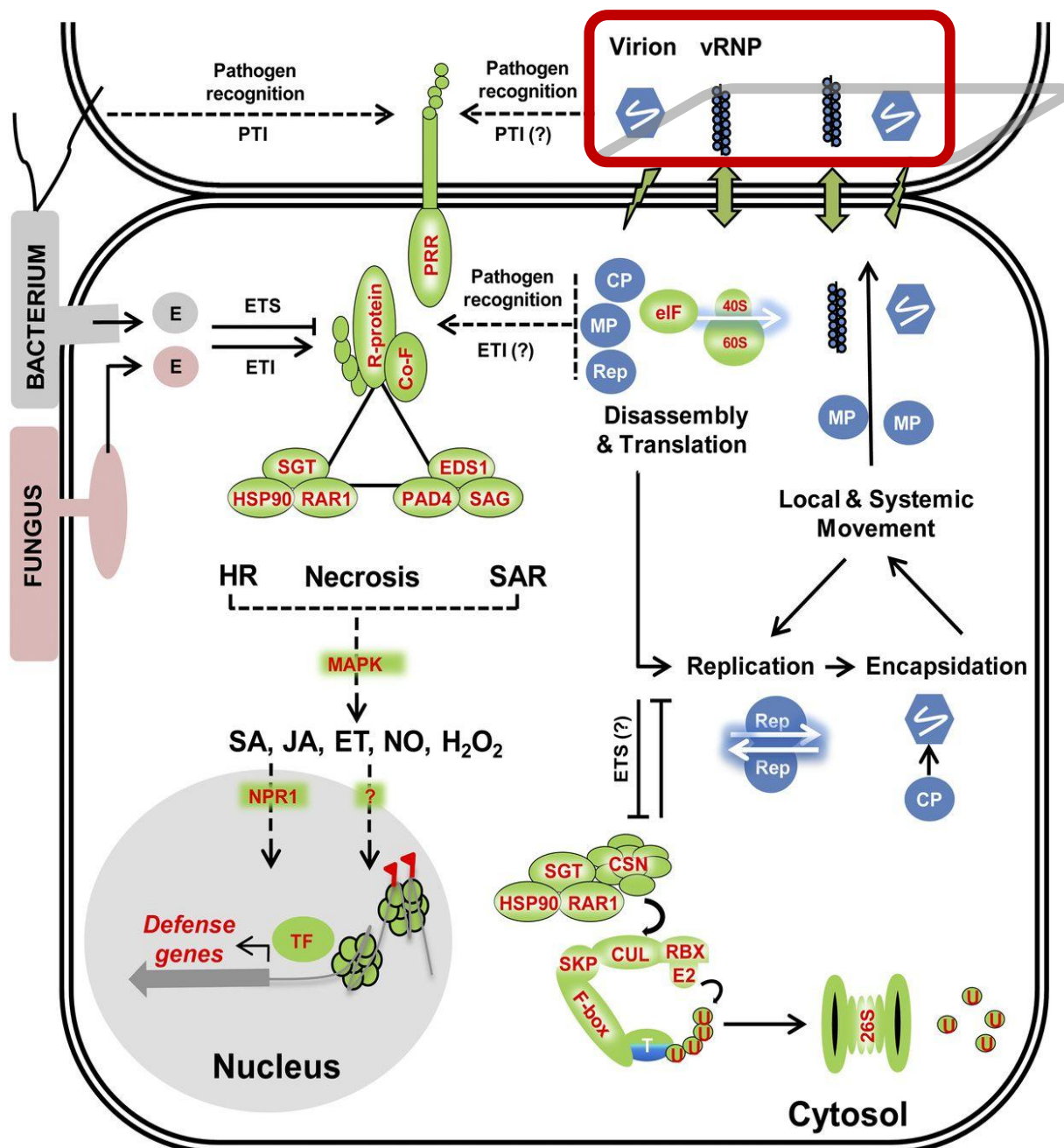


Figura 1: Respostas de defesa envolvidas na interação planta-vírus análogas às respostas de defesa contra bactérias e fungos. Os vírus entram nas plantas através de lesões celulares e se movimentam célula-célula como complexos ribonucleoproteicos e/ou virions (vRNP) através dos plasmodesmas. Proteínas codificadas por vírus como a replicase (Rep), a proteína capsidial (CP) e a proteína de movimento (MP) são traduzidas dentro do citosol da célula hospedeira. De maneira similar às infecções por bactérias e fungos, nas infecções virais, respostas imunes são disparadas quando fatores associados ao vírus são reconhecidos por putativos receptores de reconhecimento de padrões (PRRs) ou por receptores citosólicos com domínios NB-LRR. Este reconhecimento dispara respostas análogas a Imunidade ativada por efector (Effector Triggered Immunity- ETI) ou a de Suscetibilidade (Effector triggered Susceptibility- ETS) que culminam em HR, SAR ou fenótipos de necroses. Fonte: (MANDADI e SCHOLTHOF, 2013).

Embora a natureza exata da sinalização de SAR não seja evidenciada, diversos metabolitos têm sido propostos como possíveis sinalizadores de SAR em infecções virais e não viras. É provável que SAR envolva interações entre múltiplas sinalizadores como o salicilato de metila (MeSA), proteínas lipídeo-transferases e glicerol lipídeo (LIU *et al.*, 2011).

#### **1.4.4. Sistema Ubiquitina-proteosoma (UPS)**

O UPS tem emergido como um promissório sistema de defesa nas interações vírus-hospedeiro. Os vírus usam uma infinidade de estratégias para modular os processos do UPS. UPS regula atividades que incluem ciclo celular, transcrição e transdução de sinais (HERSHKO e CIECHANOVER, 1998).

O UPS nas plantas envolve primariamente a ativação da ubiquitina mediada pela enzima E1, a conjugação pela enzima E2 e a ligação pela E3 (HUA e VIERSTRA, 2011). Estas três proteínas formam o complexo E3 ubiquitina-ligase que especificamente poliubiquitina proteínas celulares que são subsequentemente alvos para a degradação pelo proteossoma 26S. SKP1 é outro componente essencial do complexo SCF (SKP1, CULLIN, e F-box) que interage com proteínas CULLIN e F-box e recruta proteínas para poliubiquitinação (HUA e VIERSTRA, 2011).

Ainda é ambíguo se os processos do UPS são usados pelas plantas para defender-se contra os vírus ou os vírus usam o UPS para promover sua virulência. Evidências das duas situações têm sido amplamente investigadas (REICHEL e BEACHY, 2000; DRUGEON e JUPIN, 2002; JIN *et al.*, 2006; LI *et al.*, 2008; ZHANG *et al.*, 2011).

Estudos para compreender o sistema UPS no hospedeiro e sua interação com as proteínas virais, bem como a identificação de proteínas alvo do UPS durante a infecção viral levaram a um maior entendimento deste processo celular associados com as respostas de imunidade antiviral e o papel crucial dos componentes do UPS à resistência ou susceptibilidade em interações vírus-hospedeiro.



### 1.5. RNA-seq no estudo de interações planta-patógeno

Dentro dos desafios que têm a área do agronegócio, está o esclarecimento dos mecanismos celulares e moleculares envolvidos na resposta de resistência ou suscetibilidade de uma planta à infecção por um determinado patógeno. Em plantas, o controle da resposta frente aos estados de estresse biótico e abiótico é mediado pela atividade transcricional de ativação e repressão de genes. A regulação da transcrição depende da ligação de ativadores ou repressores com os elementos do promotor localizados na região 5' de um gene (PROUDFOOT *et al.*, 2002). O conjunto de todos os transcritos derivados de genes produzidos numa célula em uma determinada condição fisiológica é conhecido como o transcriptoma.

A análise do transcriptoma é fundamental para compreender a função, estrutura e as interações dos genes envolvidos num determinado processo. O conhecimento global de estes mecanismos moleculares pode ser encaminhado na modulação e alteração dos padrões de expressão numa determinada condição visando melhorar e otimizar os processos biológicos envolvidos.

Na década de 1990 foram desenvolvidas algumas tecnologias para o estudo de transcriptoma como os *Northern blots*, *microarrays*, os cDNA-AFLP e análise serial de expressão de genes SAGE (*serial analysis of gene expression*). No entanto, estas tecnologias têm limitações, como o risco de hibridação cruzada, intervalo dinâmico limitado (níveis de expressão limitados) ou a necessidade de conhecer previamente o genoma de estudo (WARD *et al.*, 2012).

Em 2005 surgiu no mercado a tecnologia do RNA-seq como uma das ferramentas transcriptômicas mais promissoras, fundamentada no sequenciamento massivo de cDNA, e que se apoia no desenvolvimento das plataformas de sequenciamento de nova geração (do inglês *next generation sequencing* NGS). Atualmente, esta sendo amplamente usada em estudos de interação planta patógeno, principalmente em patossistemas que envolvem plantas modelo como *A. thaliana* (GAN *et al.*, 2011; HOWARD *et al.*, 2013) e *N. tabacum* (LU *et al.*, 2012), bem como plantas de importância econômica, como soja (KIM *et al.*, 2011; TREMBLAY *et al.*, 2011; LIN *et al.*, 2014), algodão (XU *et al.*, 2011), uva (ALABI *et al.*, 2012; PERAZZOLLI *et al.*, 2012) e maçã (GUSBERTI *et al.*, 2013). Recentemente tem-se estendido para outras

interações que envolvem plantas de menor impacto econômico, como alface (DE CREMER *et al.*, 2013) e pêssego (SOCQUET-JUGLARD *et al.*, 2013), entre outras.

O sequenciamento de RNA mediante RNA-seq tem conseguido superar as limitações de outras tecnologias de amplo uso, como os *Microarrays*, devido, principalmente, a necessidade de quantidades menores de RNA, a possibilidade de encontrar estrutura de exons, introns e locais de *splicing* alternativo, assim como permite a identificação das extremidades 5' e 3' dos genes. Além disso, mediante esta tecnologia é possível a quantificação exata dos níveis de expressão de exons e as variantes de *splicing* (MARGUERAT *et al.*, 2008; SHENDURE, 2008; WANG *et al.*, 2009).

A metodologia usada, de um modo geral consiste na purificação do mRNA, a preparação e fragmentação de uma livreria de cDNA obtida a partir do mRNA e o sequenciamento através de uma plataforma de sequenciamento (Figura 2).

No momento a tecnologia RNA-seq encontra-se disponível comercialmente em seis plataformas de NGS que estão classificadas em dois grupos. O primeiro grupo incorpora as tecnologias que estão baseadas em PCR e inclui quatro plataformas: Roche GS FLX 454 *sequencer* (Roche Diagnostics Corp., Branford, CT, USA), *Illumina genome analyzer* (Illumina Inc., San Diego, CA, USA), ABI SOLiD System (Life Technologies Corp., Carlsbad, CA, USA) e, *Ion Personal Genome Machine* (Life Technologies, South San Francisco, CA, USA). O segundo grupo inclui a HeliScope (Helicos BioScience Corp., Cambridge, MA, USA) e, *PacBio RS single-molecule real-time (SMRT) system* (Pacific Biosciences, Menlo Park, CA, USA) tecnologias baseadas no sequenciamento de uma única molécula portanto não requer o passo de amplificação prévio ao sequenciamento. Entre estas seis plataformas disponíveis o *Illumina/Solexa Genome Analyzer*, a Roche 454 GS FLX *sequencer*, o *Applied Biosystems SOLiD Analyzer* e, a HeliScope (que pertence às tecnologias de sequenciamento de segunda geração) dominam o mercado, enquanto que, a *Pacific Biosciences PacBio RS SMRT system* e, a *Ion Personal Genome Machine* da *Life Technologies* (terceira geração) têm sido introduzidas recentemente portanto ainda não são de amplo uso (JAIN *et al.*, 2014).

A escolha da plataforma mais apropriada dependerá das particularidades de um determinado projeto, exemplo: tem-se ou não um genoma de referência, a longitude das leituras que se deseja obter e a verba disponível. No caso de *Illumina* os custos são menores, no entanto as leituras que se produzem são de longitude menor em relação com a longitude das leituras geradas por Roche/454. Neste caso é necessário um novo sequenciamento, leituras maiores são recomendáveis (OZSOLAK e MILOS, 2011; BARBA *et al.*, 2014).

A Tabela 1 mostra um resumo comparativo entre as principais plataformas de sequenciamento baseado em (BARBA *et al.*, 2014), o qual facilita a escolha da plataforma mais apropriada. Em resumidas contas, Roche 454 gera as leituras de maior longitude *Illumina* tem a maior capacidade de sequenciamento e os menores custos e, SOLiD 5500 xls a maior acurácia (LIU *et al.*, 2012).

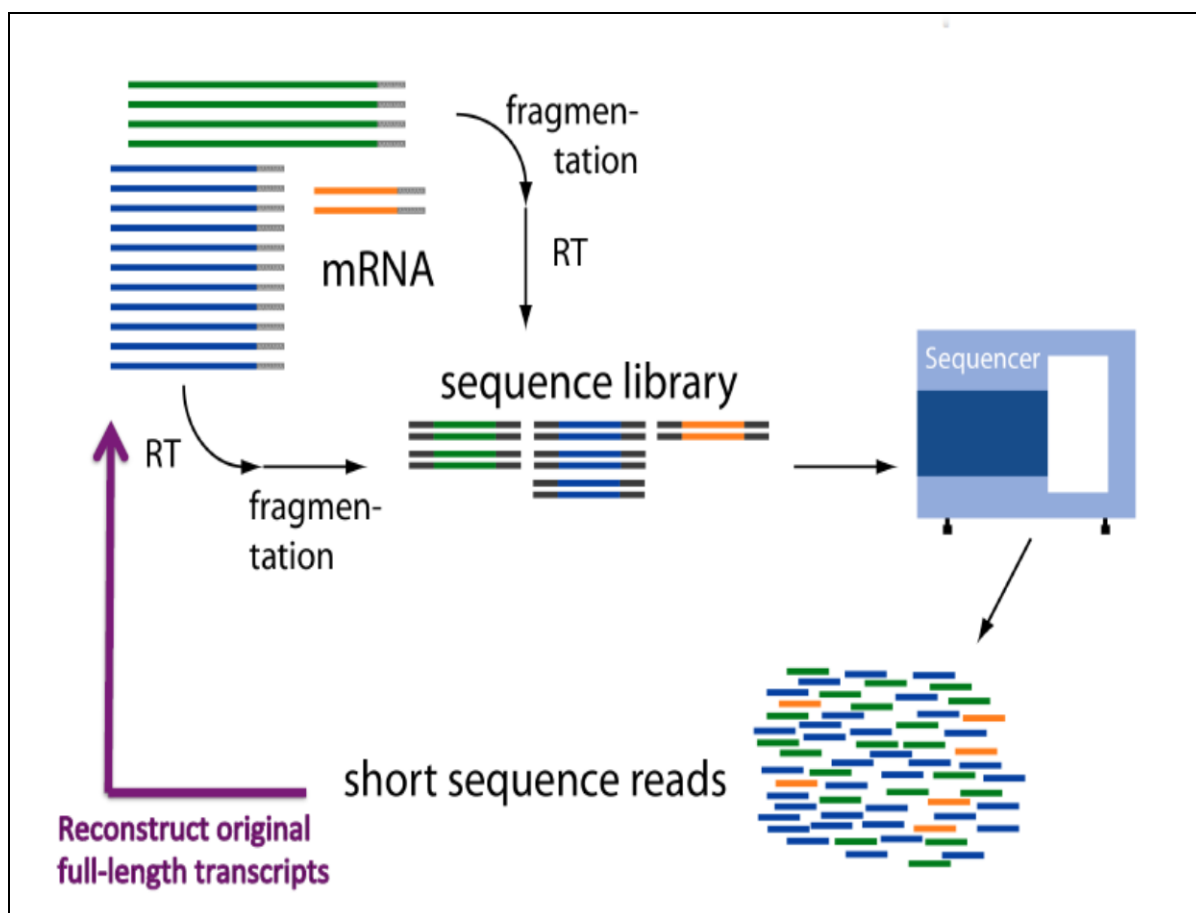


Figura 2: Metodologia geral usada na técnica de sequenciamento RNA-seq. Uma livreria de cDNA deve ser preparada após do isolamento e fragmentação do mRNA. Esta livreria será sequenciada usando uma plataforma de sequenciamento que gera milhões de leituras cortas. Fonte: (MSKCC, 2014)

Como foi exposto anteriormente, cada plataforma de sequenciamento tem suas vantagens e suas desvantagens. Porém, para projetos de *RNA sequencing*, é preciso ter uma alta cobertura (do inglês *high depths of coverage*: uma media alta de leituras que estejam sobrepondo num determinado nucleotídeo na sequência reconstruída). A *Illumina* é uma das opções mais recomendáveis para este tipo de projetos a qual oferece o menor custo, uma boa precisão, e o maior rendimento (RADFORD *et al.*, 2012) e, que durante os últimos 5 anos, tem sido usada com maior frequência em diferentes projetos que envolvem sequenciamento em virologia de plantas (BARBA *et al.*, 2014).

A plataforma de sequenciamento escolhida para este projeto foi a plataforma *Illumina Hiseq 2000*.

A *Illumina* nos últimos 3-4 anos tem desenvolvido a serie de plataformas que incluem a HiSeq® 2500, a HiSeq 2000, a HiSeq 1500 e a HiSeq 1000 as quais têm vindo tomando vantagem em relação com outras plataformas, devido principalmente à quantidade de nucleotídeos que são capazes de sequenciar numa mesma corrida ou no inglês *one lane*, bem como, o tempo que leva o sequenciamento, a longitude das leituras geradas, a precisão no sequenciamento e os baixos custos.

A *Hiseq 2500* tem a capacidade de sequenciar um genoma em 24 horas, 20 exomes num dia ou 30 amostras para *RNA sequencing* em aproximadamente 5 horas. A *Hiseq 2000* é capaz de sequenciar 600 bilhões de pb per corrida, enquanto que outras plataformas como a *SOLiD 5500 xlw da Applied Biosystem* gera 30 bilhões de pb por corrida com leituras de 85 pb de longitude mas com uma excelente precisão do 99,99%, por sua parte *454 GS-FLX+ Titanium* da Roche é capaz de gerar aproximadamente 600 milhões de pb por corrida que é bastante menor que as *Hiseq*, mas que produz leituras com uma longitude que podem alcançar até os 1000 pb, enquanto que a *Hiseq 2500* somente gera leituras de máximo 200 pb (BARBA *et al.*, 2014).

O verdadeiro desafio dentro de um projeto de RNA-seq consiste na analises compreensão e interpretação da grande quantidade de dados gerados cujo objetivo é reconstruir o transcriptoma a partir das milhões de leituras e, após encontrar padrões que respondam uma pergunta biológica. O processamento das leituras

requer de um controle de qualidade da sequência das leituras e após, dependendo do objetivo do experimento, as leituras geradas a partir das diferentes livrarias de cDNA podem ser montadas mediante um de *novu assembly* quando não se têm um genoma de referência ou alinhadas usando um genoma de referência.

A montagem, contagem, normalização e análises estatísticos requeridos para o processamento das gigas pares de bases (Gpb) de informação produzidas no sequenciamento é feito com programas bioinformáticos especializados que têm se desenvolvido de maneira paralela ao avanço das NGS.

Atualmente estão disponíveis muitas ferramentas para este tipo de análises tanto de uso livre (*open source*) como de uso comercial. Consequentemente uma compreensão completa é necessária para escolha da *pipeline* informática mais apropriada, do mesmo modo que pessoas com habilidades em bioinformática são requeridas para desenvolver uma adequada e robusta análise de dados a fim de maximizar as informações obtidas.

### **1.6. Bioinformática no processamento de dados gerados a partir da *Illumina***

Para o processamento bioinformático de dados, embora que já existam programas comerciais como *Geneious* ou *CLC Bio*, bem como programas de uso livre como *Galaxy*, que possuem interfaces gráficas e, que permitem ao usuário fazer análises básicas de um modo simples, quando grandes quantidades de dados devem ser processados e resultados com maior acurácia são desejados, é recomendável usar ferramentas que estejam baseadas em algoritmos robustos, eficientes e com princípios estatísticos.

Trapnell *et al.* (2012) desenvolveram uma *pipeline* informática para dados sequenciados em *Illumina* ou *SOLiD*, que já possuem um genoma de referência. Esta *pipeline*, está formada por duas ferramentas, *TopHat* e *Cufflinks*, que em conjunto resolvem 3 das etapas essenciais numa análise de dados de RNA-seq: (i) alinhamento; (ii) montagem de transcritos ou anotação do genoma e, (iii) quantificação de genes e transcritos.

*TopHat* alinha as leituras ao genoma e descobre *splice sites*. *Cufflinks* usa este mapa contra o genoma para montar as leituras nos transcritos. *Cuffdiff* como parte de *Cufflinks*, toma as leituras alinhadas a partir de dois ou mais condições e reporta os genes e transcritos que são diferencialmente expressos usando uma rigorosa análise estatística. *TopHat* e *Cufflinks*, têm uma ampla aceitação e, vem sendo utilizados em recentes estudos transcriptômicos de alta resolução (GRAVELEY *et al.*, 2010; MIZUNO *et al.*, 2010; LISTER *et al.*, 2011; TWINE *et al.*, 2011).

*TopHat* é uma das ferramentas de uso livre de maior confiabilidade, eficiência e aceitação para o alinhamento de leituras geradas a partir de plataformas de sequenciamento de nova geração usando os *supercontigs* como referência sem conhecimento prévio de *splice sites* (TRAPNELL *et al.*, 2009).

O alinhamento e montagem das leituras foi feita usando como referência o genoma de *C. papaya* (MING *et al.*, 2008) que foi publicado na revista *Nature*, possui um genoma de 370 Mb organizados em *contigs* e *scaffolds* contendo um total de 24.746 genes e que foi gerado a partir de uma planta feminina do cultivar transgênico SunUp. Porém, o genoma completo ainda não está disponível (somente *contigs* e *supercontigs* disponíveis), fator que é importante de considerar porque entre maior seja o nível de compactação da informação, o trabalho torna-se menos complexo bem como, maiores ferramentas de análises se encontram disponíveis.

Para a contagem de leituras existem dois principais métodos, o primeiro usa o número de leituras por kilobase por milhão de leituras mapeadas (do inglês *Reads Per Kilobase per Million mapped reads* RPKM), método que reflete a concentração molar de um transcrito na amostra inicial mediante a normalização por longitude do RNA, bem como pelo número total de leituras na contagem (MORTAZAVI *et al.*, 2008).

Este tipo de normalização facilita uma comparação nos níveis de transcrito dentro e entre as amostras, porém não leva em consideração que os experimentos de RNA-seq *paired-end* produzem duas leituras por fragmento, mas não necessariamente as duas leituras serão mapeáveis, por exemplo, a segunda leitura pode ter uma baixa qualidade e serem excluída. Se fossem contadas as leituras na vez de fragmentos, alguns fragmentos poderiam ter uma dupla contagem, enquanto que outros não,

levando a contagens enviesadas. Portanto *Cufflinks* usa o método de contagem fragmentos por kilobase de transcrito per milhão de fragmentos mapeados (FPKM) que é uma alternativa que soluciona as falhas da contagem por RPKM (TRAPNELL *et al.*, 2010).

Para encontrar os genes diferencialmente expressos o programa *cuffdiff* 2 incluso no *Cufflinks* assume que a expressão de um transcrito em cada condição pode ser medida pela contagem do número de fragmentos gerados a partir de este. Portanto a alteração nos níveis de expressão de um transcrito é medido pela comparação da contagem dos fragmentos em cada condição.

*Cuffdiff* usa por defeito uma relação  $\log_2$  (*fold change*) como critério para considerar um cambio como significativo entre a contagem de fragmentos para um gene determinado entre duas condições. A razão  $\log_2$  fold change é dada pela expressão:  $\log_2(\text{amostra 2}/\text{amostra 1})$  onde a amostra 2 e a condição que esta sendo avaliada em relação com a amostra 1 que é a condição de referência. *Cuffdiff* faz um teste estatístico para identificar os genes e transcritos diferencialmente expressos no qual o *log-fold change* observado na expressão dos genes é avaliado contra a hipóteses nula de não cambio (o verdadeiro *log-Fold change* é zero).

A significância estatística usando um modelo de variabilidade no *log-fold change* submetido à hipóteses nula, se faz necessária, levando em consideração que os erros na medição, a variabilidade técnica e a variabilidade biológica a través das réplicas poderiam levar a que um observado *Log-fold change* fosse diferente de zero.

Este modelo descrito em detalhe no (TRAPNELL *et al.*, 2013) procura controlar adequadamente os pontos críticos do planejamento experimental descritos anteriormente, bem como eventos próprios do mecanismo transcripcional como o *splicing* alternativo, pontos que outros modelos não conseguem resolver em conjunto. Consequentemente, *Cuffdiff* para resolver estas questões faz a modelagem da variabilidade na contagem de fragmentos dos transcritos em função de sua expressão e de sua estrutura de *splicing*.

Primeiro, *Cuffdiff* determina o grau de sobredispersão, ajustando a variância na

contagem dos fragmentos em função de uma média obtida das réplicas experimentais. Segundo, estima a incerteza (do inglês *uncertainly*) calculando o grau de confiança de que um determinado fragmento seja atribuído corretamente ao transcrito do qual foi gerado. Transcritos com mais éxons compartilhados e com alguns fragmentos exclusivamente atribuídos, terão um maior grau de incerteza. Finalmente, combina a incerteza obtida em cada contagem dos fragmentos dos transcritos com a sobredispersão prevista para cada contagem.

A incerteza é calculada mediante um algoritmo que interpreta a contagem de fragmentos para um transcrito, como uma distribuição beta e, a sobredispersão na contagem, como uma distribuição binomial negativa. O mesmo algoritmo mistura as duas distribuições, para interpretar tudo como uma distribuição beta binomial negativa. Finalmente a alteração na expressão de cada gene e transcrito entre duas ou mais condições é reportada com sua correspondente significância estatística (valor p) ajustados ao método estadístico *false discovery rate (FDR)* (BENJAMINI e HOCHBERG, 1995). Este método se faz necessário quando se tem medições de milhares de variáveis a partir de um pequeno grupo de réplicas. Caso o valor p for 0,05 se aceita que 5% de todos os testes serão falsos positivos. Entretanto se o valor p for ajustado a uma FDR de 0,05 se aceita que 5% dos testes encontrados estatisticamente significativos serão falsos positivos.

Para encerrar, de modo geral poucos trabalhos que abordem o estudo da expressão de genes envolvida *C. papaya*, têm sido feitos e segundo a recopilção que faz (TRIPATHI *et al.*, 2014) dos trabalhos que envolvem estudos de transcriptômica em mamão apenas um deles (ARYAL *et al.*, 2012) aborda interações planta hospedeiro.

Visando entender os mecanismos moleculares envolvidos no aparecimento dos sintomas e, a resposta do mamoeiro ao PMeV, um estudo transcriptômico foi abordado. Neste estudo nos foi analisada a alteração no transcriptoma de mamão em resposta à inoculação com o PMeV em três pontos chave no aparecimento dos sintomas em relação com plantas que se encontravam no estágio de prefloração.



Tabela 1. Resumo comparativo entre as principais plataformas de sequenciamento. Fonte: Adaptada de: Barba, 2014.

Plataforma	Método de amplificação	Química do sequenciamento	Comprimento das leituras (pb)	Máxima produção per corrida	Acurácia (%)
454 (Roche)	PCR de emulsão	Pirosequenciamento	400-700	700 Mpb	99.9
<i>Illumina</i> (ILLUMINA)	Amplificação em ponte ( <i>Bridge PCR</i> )	Terminadores reversíveis	100-300	600 Gpb	99.9
SOLiD (Life Technologies)	PCR de emulsão	Ligação	75-85	80-360 Gpb	99.99
PacBio (Pacific Biosciences)	Sequenciamento de molécula única em tempo real	Nucleotídeos fluorescentemente marcados	4000-5000	200 Mb-1 Gb	95
Helicos (Helicos Biosciences)	Sequenciamento de molécula única	Terminadores reversíveis	25-55	35 Gpb	97
Ion Torrent (Life Technologies)	PCR de emulsão	Detecção da liberação do H <sup>+</sup>	100-400	100 Mb-64Gpb	99
Nanopore (Oxford Technologies)	Sequenciamento de molécula única	-	Leituras muito extensas até de 50 kpb	Dezenas de Gpb	96

## **2. OBJETIVOS:**

### **2.1 Objetivo geral:**

Identificar os genes diferencialmente expressos durante a interação mamoeiro-PMeV e indicar alguns genes chaves na indução a resistência ao PMeV.

### **2.2 Objetivos específicos:**

- Delinear o planejamento experimental para o sequenciamento por RNA-seq;
- Preparar e avaliar a qualidade das amostras a serem sequenciadas;
- Analisar e processar os dados produzidos no sequenciamento;
- Encontrar padrões de expressão gênica entre os genes diferencialmente expressos;
- Postular genes que possam estar envolvidos na resistência do mamão o PMeV.

### 3. MATERIAIS E MÉTODOS

Este projeto envolveu três etapas gerais: uma etapa *in vivo*, que compreende o processamento das amostras para o sequenciamento, o sequenciamento e as análises *in silico*, que agrupam os métodos usados no processamento e análises de dados (Figura 3).

#### 3.1. Material Vegetal

Mudas do cultivar Golden de *Carica papaya* de aproximadamente 30 dias foram plantadas em campo na fazenda experimental do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) localizada no município de Sooretama-ES, Brasil.

#### 3.2. Inoculação das plantas de mamoeiro

A inoculação foi feita no ápice do caule de plantas hermafroditas com 2 meses de idade (pós-plantio) com auxílio de uma seringa. Dentre as plantas cultivadas em campo, 03 plantas foram inoculadas com 20 µl de látex infectado com PMeV diluído em tampão fosfato de sódio 50 Mm pH 7,0 (1:1), grupo de plantas inoculadas com o PMeV (I) e outras 03 plantas foram inoculadas somente com 20 µl de tampão fosfato de sódio 50 Mm pH 7,0, grupo de plantas não inoculadas com o PMeV (NI).

#### 3.3. Coleta e armazenamento

Ao longo do experimento em campo, amostras de folhas de mamoeiro com 2, 3, 6 e 8 meses de idade foram coletadas. A Figura 4 (a-i) mostra as plantas inoculadas com PMeV e a Figura 5 (a-i), as plantas não inoculadas com PMeV. As amostras foram imediatamente congeladas em nitrogênio líquido para o transporte até o laboratório e armazenadas a -80 °C até o isolamento do RNA.

Durante a coleta, a presença ou a ausência dos sintomas da meleira do mamoeiro

foram anotadas, conforme apresentado na Tabela 2. Em todos os casos, a coleta do material vegetal foi feito a partir de uma das folhas mais novas.

Tabela 2: Observações feitas das plantas em cada ponto avaliado.

Idade (meses)	Observações
2	Idade em que foi feita a inoculação. Plantas em estágio de prefloração, sem sintomas.
3	Plantas com botão floral, sem sintomas.
6	Plantas com fruto, mas ainda sem sintomas.
8	Plantas com fruto no começo da maturação. Plantas com sintomas de meleira.

### 3. 4. Delineamento experimental

O delineamento experimental seguido encontra-se resumido na Tabela 3. Foram avaliados dois tratamentos: plantas não inoculadas (NI) com três repetições biológicas (P03, P08, P33) e plantas inoculadas (I) com três repetições biológicas (P02, P16, P22), em 3 intervalos de tempo: 3, 6 e 8 meses de idade. Para as análises de expressão diferencial de genes, os 3 pontos foram comparados em relação com as plantas de 2 meses de idade que não se encontravam infectadas pelo PMeV e que estavam em prefloração.

Tabela 3: Delineamento experimental

2 m	Idade (meses)	3	6	8
	Tratamento			
Planta 03	T1 (NI)	Planta 03		
Planta 08		Planta 08		
Planta 33		Planta 33		
Planta 02	T2 (I)	Planta 02		
Planta 16		Planta 16		
Planta 22		Planta 22		

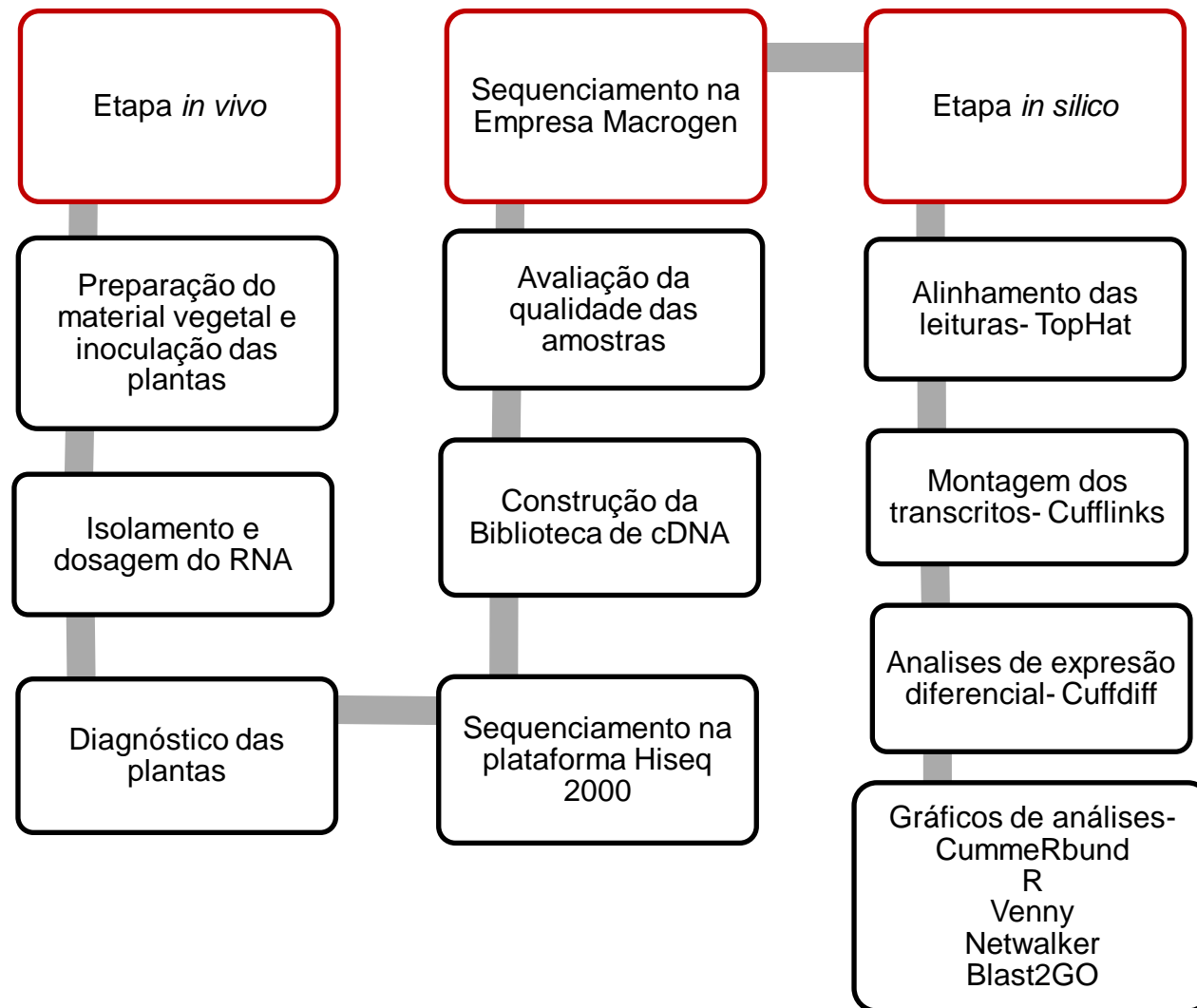


Figura 3: Fluxograma seguido no desenvolvimento da metodologia agrupado em três etapas principais mostradas em cor vermelho.



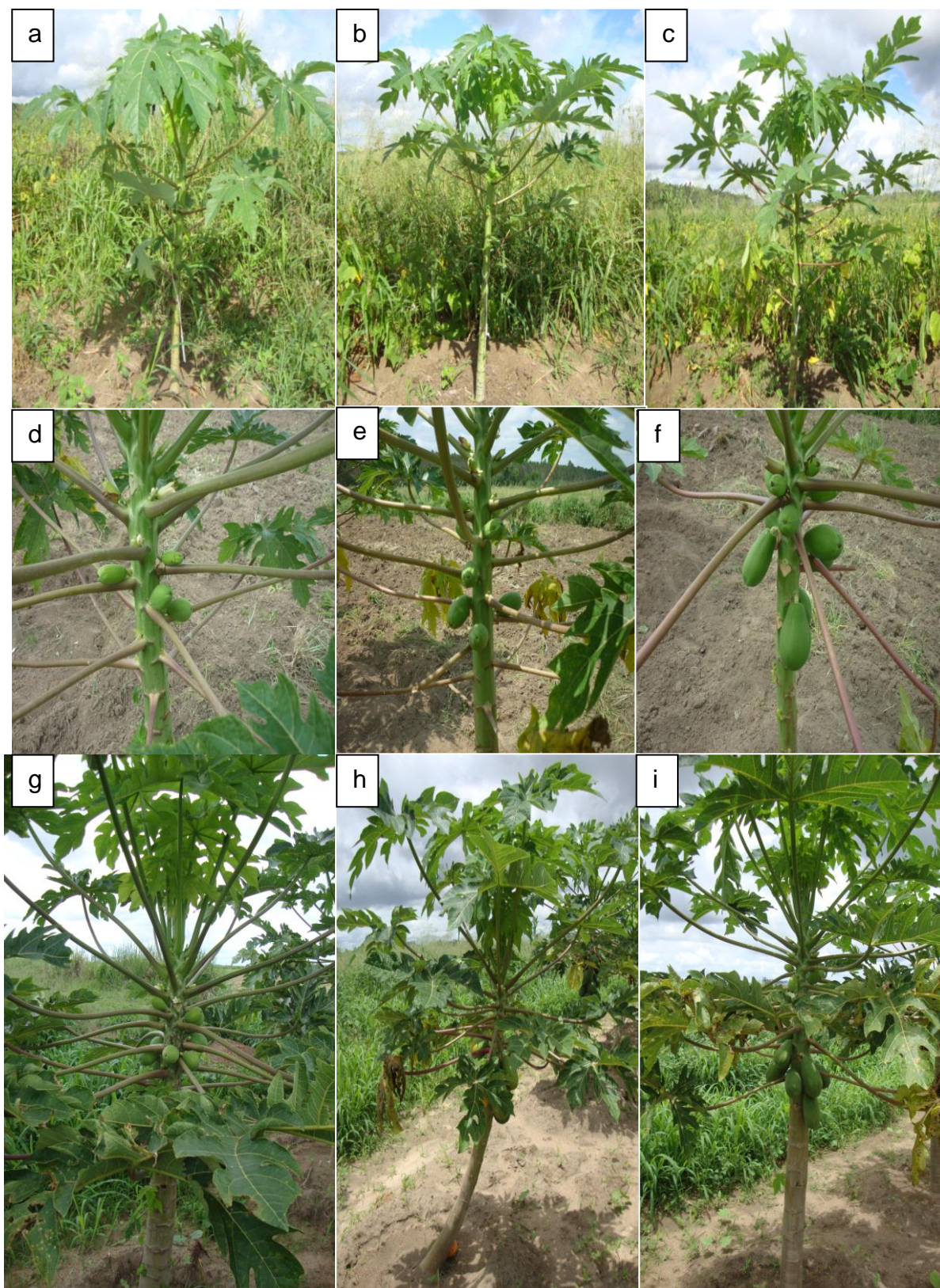


Figura 4: Plantas Inoculadas com látex infectado com PMeV diluído em tampão fosfato de sódio 50Mm pH 7,0 (1:1). (a, b, c) representam plantas com 3 meses de idade, (d, e, f), com 6 meses de idade e (g, h, i), com 8 meses de idade.



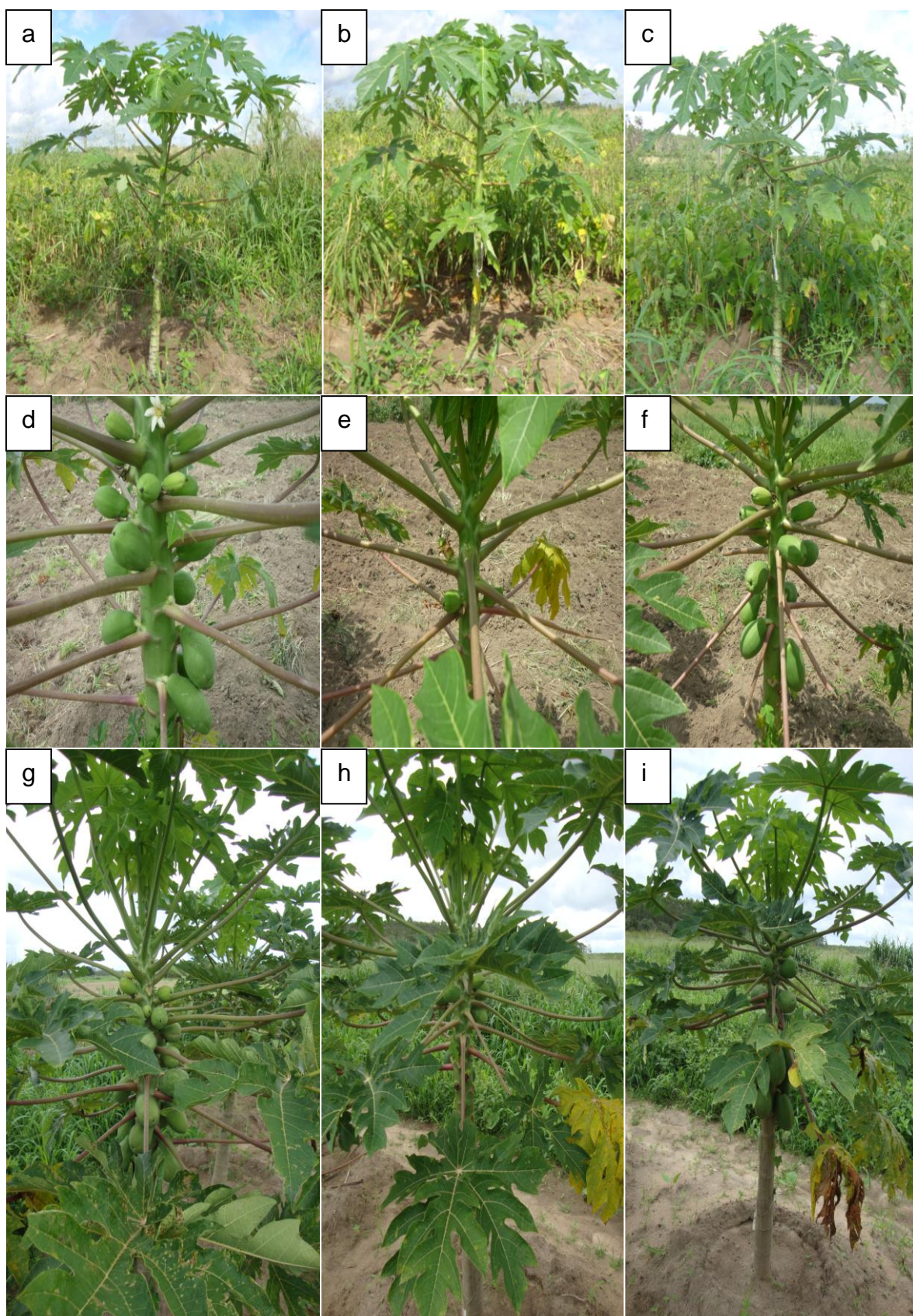


Figura 5: Plantas inoculadas com tampão fosfato de sódio 50 Mm pH 7.0. (a, b, c) representam plantas com 3 meses de idade, (d, e, f), com 6 meses de idade e (g, h, i), com 8 meses de idade.

### 3.4. Extração de RNA e diagnóstico molecular do PMeV:

A extração de RNA foi realizada a partir de 100 mg de folha. O protocolo usado foi o indicado no *RNAeasy plant mini kit* (Qiagen Inc., EUA cat. 74904). Uma alíquota de cada amostra foi usada para fazer o diagnóstico do PMeV, posteriores análises e validação dos resultados mediante outras técnicas.

O diagnóstico molecular de meleira mediante RT-PCR convencional foi realizado seguindo o protocolo descrito em (ABREU *et al.*, 2012). O RNA foi dosado usando o espectrofotômetro *NanoDrop 2000* da *Thermo Scientific* (EUA) e posteriormente tratado com uma solução estabilizadora de RNA seguindo o protocolo indicado para o uso do produto *RNAstable solution* (Sigma-Aldrich, St. Louis, MO, EUA).

Finalmente, 24 amostras contendo 30 µl de RNA com concentrações superiores a 4 µg foram enviadas para seu sequenciamento em MacroGen Inc. (Korea), sendo a qualidade das amostras prévio ao sequenciamento avaliado pela empresa. O procedimento usado é brevemente descrito a seguir.

### 3.5. Testes de qualidade do RNA:

Antes de proceder na construção da biblioteca de cDNA, a empresa MacroGen avalia a qualidade das amostras usando o *Bioanalyzer 2100* da *Agilent Technologies* mediante o método *RNA integrity number* (RIN) que estima a integridade do RNA. Amostras com um RIN (próximo ou maior do que 7,0), ou razão perto de 1,5 e uma quantidade próxima aos 4 µg são recomendadas para prosseguir com a construção da biblioteca.

Como apresentado na Tabela 4, embora nem sempre os dois parâmetros para medir a integridade do RNA foram os desejados, na maioria das amostras a quantidade de RNA obtida foi bastante alta e suficiente para manter delineamento experimental e continuar com a preparação da biblioteca de cDNA.



Tabela 4: Controle de qualidade do RNA.

Amostra	Quantidade total (ug)	RIN	rRNA Relação
P03C1	29.1	6.4	0.7
P08C1	26.6	7.4	1.2
P33C1	22.6	6.9	1.1
P02C1	28.2	7.1	1.2
P16C1	32.2	7.4	1.2
P22C1	38.0	7.2	1.2
P03C3	23.2	2.9	0.7
P08C3	18.3	7.6	1.4
P33C3	21.6	7.3	1.2
P02C3	15.8	7.4	1.3
P16C3	4.7	7.3	1.3
P22C3	8.3	6.9	0.9
P03C7	20.2	6.1	1.1
P08C7	20.2	6	1.4
P33C7	11.1	8	1.5
P02C7	21.7	7.5	1.4
P16C7	26.9	5.8	0.5
P22C7	7.8	7.8	1.7
P03C11	5.9	6.2	1.1
P08C11	3.7	7.5	1.8
P33C11	5.6	7.4	1.5
P02C11	7.5	7.3	1.3
P16C11	4.8	7.3	1.5
P22C11	3.0	7	1.2

### 3.6. Construção da Biblioteca de cDNA.

Para a construção da biblioteca a partir das amostras de *C. papaya*, a MacroGen utilizou a metodologia descrita como *TruSeq mRNA library construction*, onde fragmentos com um tamanho entre as 200-400 pb e concentrações superiores a 1,5 ng/μl são requeridos para proceder com o sequenciamento.

A preparação da biblioteca compreende passos que iniciam com o Isolamento e fragmentação do mRNA, continuam com a síntese do cDNA, para o qual se usam hexâmeros randômicos e transcriptase reversa. Após, modificações requeridas para a amplificação dos fragmentos de cDNA mediante uma PCR em ponte (do inglês *Bridge PCR*) previa ao sequenciamento, são feitas. Estas modificações incluem a adição de uma Adenina nos extremos 3' dos fragmentos de cDNA aonde sequencias curtas de DNA dupla fita chamadas de adaptadores, do inglês *adapters*, são ligados por complementaridade de bases. Finalmente os fragmentos que possuem os adaptadores em ambas extremidades são ligados a *templates* complementares que estão fixos numa superfície sólida chamada célula de fluxo do inglês *flow cell* (Figura 6).

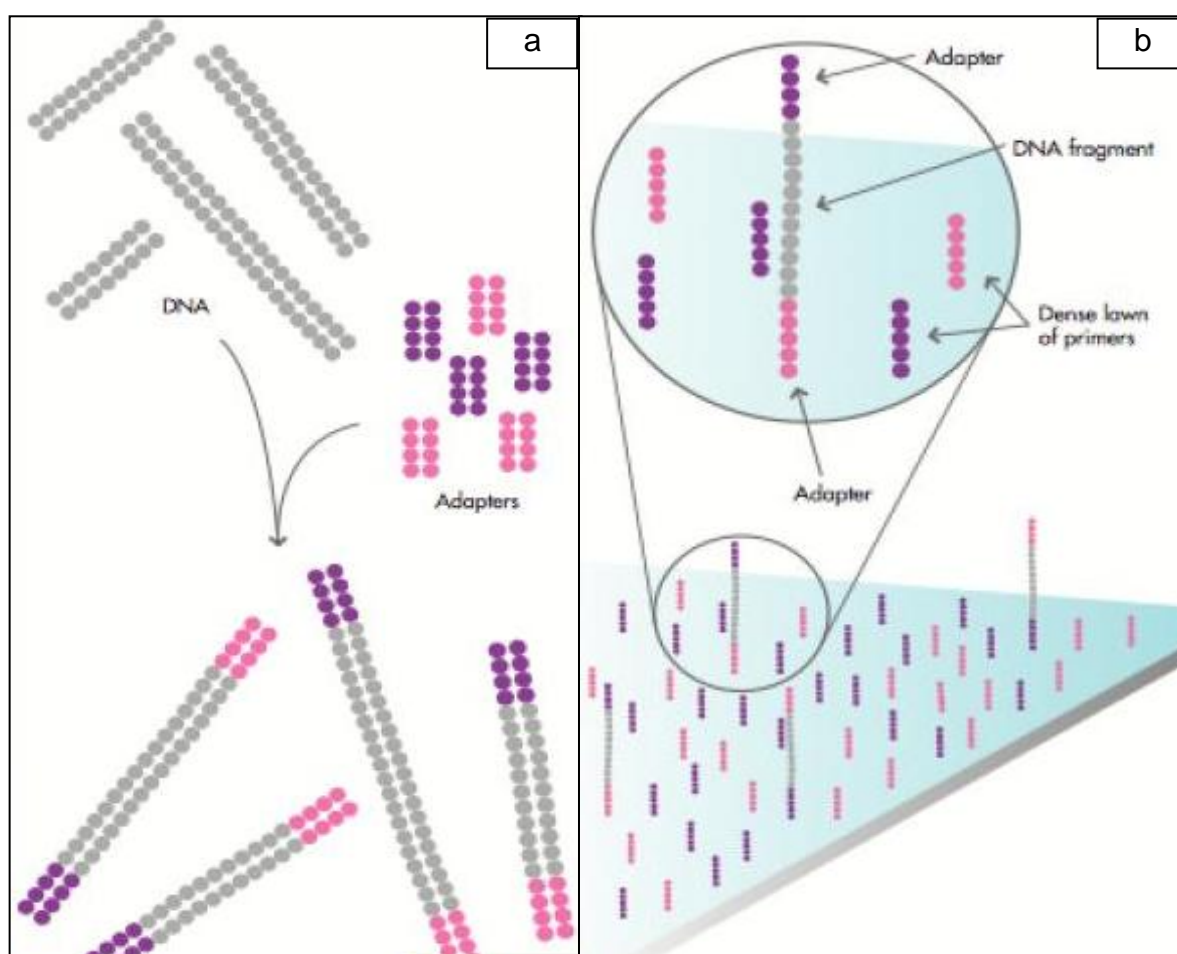


Figura 6: (a) Ligação das seqüências adaptadores aos fragmentos de DNA obtidos a partir da construção da biblioteca de cDNA e posteriores modificações. (b) Ligação dos fragmentos de DNA+ as seqüências adaptadores com os templates complementares fixos na célula de fluxo sólida. Fonte: adaptado de (ILLUMINA, 2010).

Usando esta metodologia foram obtidos fragmentos de cDNA que oscilaram entre as 263 e 292 pb e, as concentrações entre os 39.33 e 76.27 ng/μl (apêndice A).

### 3.8. Sequenciamento

Para o sequenciamento dos fragmentos do cDNA, a empresa MacroGen usou a tecnologia da plataforma *Illumina* HiSeq 2000. Prévio ao sequenciamento, uma reação de amplificação em ponte é feita numa célula de fluxo ou *flow cell*. Os *templates* ligados aos adaptadores dos fragmentos de cDNA agem como iniciadores *senso-antisense* originando pontes que favorecem a amplificação na presença de nucleotídeos não marcados e da enzima polimerase. Os amplicons ficam aderidos e, depois de uma desnaturação, formarão uma nova ponte para permitir a amplificação. Estes passos são repetidos sucessivamente, gerando-se milhões de grupos ou *clusters* de um determinado fragmento (Figura 7).

A *flow cell* contendo os clusters já formados é colocada no HiSeq 2000, onde uma nova desnaturação é feita para dar início a ciclos automatizados de extensão e captura de imagem. Nesta ocasião, nucleotídeos marcados com fluoróforos reversíveis são introduzidos na reação. Estes nucleotídeos têm propriedades de terminação, o que permite parar a síntese de DNA quando a DNA polimerase integra o correspondente nucleotídeo na fita nascente.

Uma vez integrado o nucleotídeo na fita nascente, os fluoróforos são ativados por um laser. A luz emitida será diferencial dependendo do nucleotídeo incorporado. Esta informação é capturada e armazenada. Uma vez terminado o processo anterior, os nucleotídeos não integrados na reação são retirados e, enzimaticamente, é cortado o "terminador" para que um novo ciclo permita a incorporação do seguinte nucleotídeo.

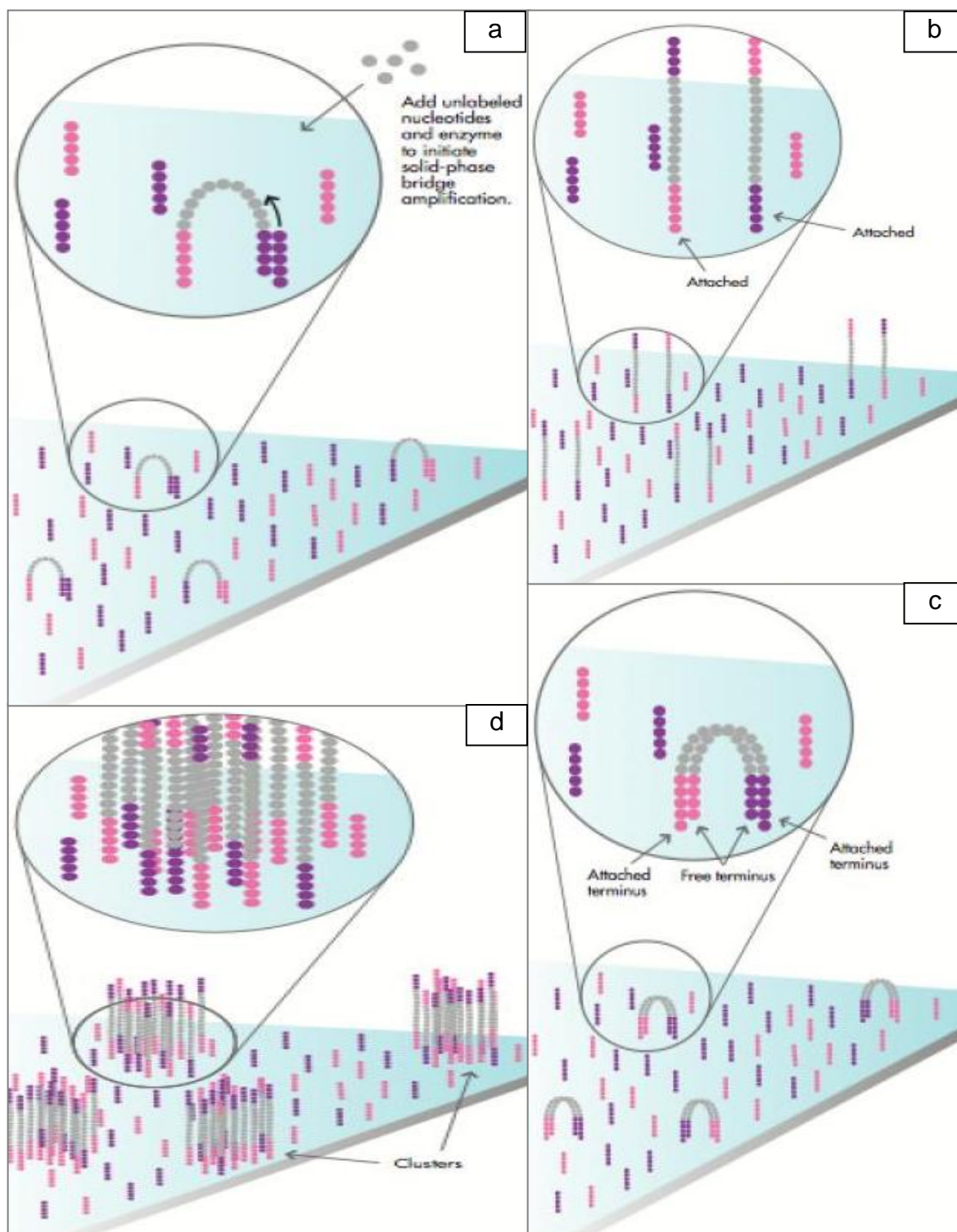


Figura 7: Amplificação em ponte. (a) Os *templates* ligados nos adaptadores dos fragmentos de cDNA ou DNA agem como iniciadores formando pontes para favorecer a amplificação em presença da enzima polimerase e de nucleotídeos na marcados.(b) Os amplicons são desnaturados a desnaturação e (c) posteriormente formarão uma nova ponte. Os ciclos são repetidos sucessivamente formando (d) *clusters* para o sequenciamento. Fonte: adaptado de (ILLUMINA, 2010).

### 3.9. Análises de bioinformática

A análise *in silico* foi feita usando programas bioinformáticos de código aberto (do inglês: *open source*).

#### 3.9.1. Verificação da qualidade das leituras geradas

A verificação da qualidade das leituras geradas (raw sequence) foi feita na empresa MacroGen usando o software FastQC v0.10.0. Parâmetros como a qualidade da sequência por nucleotídeo (Q20 e Q30), por conteúdo de GC, por conteúdo de nucleotídeos não determinados (N) e outros são avaliados com este software. FastQ usa o phred quality score para determinar os Q20 e Q30 score. Este é definido pela seguinte expressão:

$$Q = -10 \log_{10} P$$

Em que  $P$  é a probabilidade de que uma base na leitura tenha sido atribuída de modo errado durante o sequenciamento.

#### 3.9.2 Alinhamento das leituras, montagem de transcritos e análises de expressão diferencial de genes.

Após da avaliação da qualidade das leituras estas foram processadas usando a *pipeline* bioinformática *TopHat-Cufflinks-cuffdiff-cummeRbund* do jeito que é exemplificado no apêndice B.

Para o alinhamento das leituras foi usado o software *TopHat* v1.3.3, disponível no site <http://ccb.jhu.edu/software/tophat/index.shtml>. *TopHat* é um mapeador *splice junction* que usa o *Bowtie aligner* para alinhar as leituras ao genoma de referência e posteriormente identifica *splice junction* entre os exons. O genoma de referência usado foi o de *Carica papaya* v9.0 disponível em <http://www.phytozome.net/papaya.php>.

Na montagem dos transcritos, determinação de abundancias e análises de

expressão diferencial foi usado o software *Cufflinks* v2.2.1 disponível no site: <http://cufflinks.cbc.umd.edu/>.

A análise de expressão diferencial de genes (*Differentially expression analysis* - DEG) foi feita usando o programa *cuffdiff* incluso no pacote de *cufflinks*. Foram considerados como diferencialmente expressos apenas os genes cujos valores *p* foram ajustados a uma FDR de 0.05.

### 3.10. Análises de dados e gráficos de expressão

Nas análises dos dados obtidos foram utilizados vários programas que são apresentados a continuação, algumas ferramentas de análises disponíveis *online*, bem como códigos ou *scripts* para resolver tarefas simples que necessitaram ser implementados na linguagem de programação de *linux* e *python* (módulo *Biopython*).

Na visualização gráfica da distribuição geral dos dados em relação aos níveis de expressão, o programa R v3.1.1 disponível em <http://www.r-project.org/> foi usado para a construção do diagrama de caixa e diagrama de densidade de dados. Para as análises de correlação entre as repetições biológicas foi usado o módulo de R *cummeRbund*.

As relações entre os conjuntos de genes expressos diferencialmente nas duas condições, nos três tempos de desenvolvimento fisiológico, foram analisadas por meio de diagramas de *venn*, usando a ferramenta interativa online, *Venny*, desenvolvida por (OLIVEROS, 2007). Os elementos comuns identificados nos diagramas de *venn* foram agrupados por sua similaridade relacionada com os níveis de expressão gênica usando o programa *NetWalker* v1.0. disponível em <https://netwalkersuite.org/tutorials/doxorubicin/clustering-heatmap-analysis>, o qual fez o agrupamento hierárquico por distância euclidiana com o método UPGMA.

Dentro dos clusters nos quais se encontrou um padrão diferencial de expressão relacionado com a inoculação do vírus, os genes foram selecionados e sua anotação funcional foi feita usando o programa *Blast2GO* v2.7.1. disponível em <https://www.blast2go.com/b2ghome>.

## 4. RESULTADOS E DISCUSSÃO

### 4.1. Diagnóstico do PMeV nas amostras selecionadas para o sequenciamento

O resultado do diagnóstico realizado utilizando a técnica de RT-PCR nas seis plantas selecionadas para o RNA-seq aos 2, 3, 6 e 8 meses é apresentado na Figura 8. Na primeira coleta, correspondente a plantas com 2 meses de idade, todas as amostras foram negativas para a infecção do vírus, conforme previsto pois o material analisado foi procedente de folhas que ainda não haviam sido inoculadas. De modo similar, as plantas que não foram inoculadas com o PMeV resultaram negativas em todos os tempos avaliados.

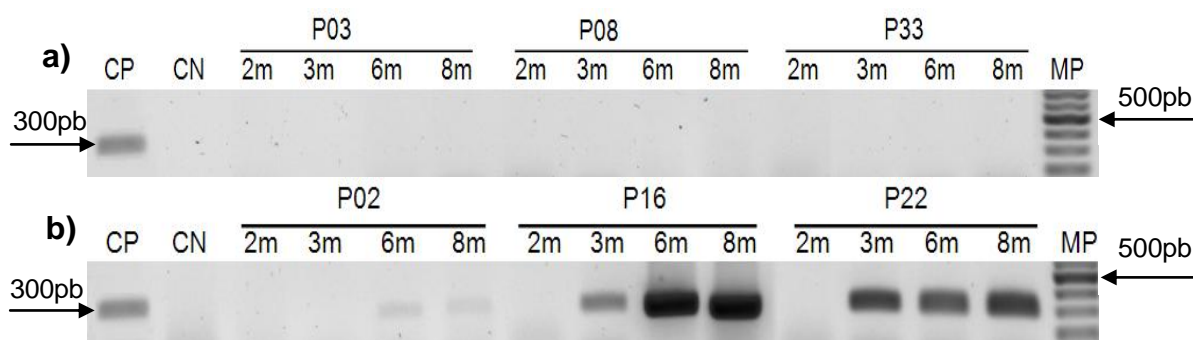


Figura 8: Diagnóstico de meleira. (a) plantas não inoculadas e (b) plantas inoculadas com PMeV. CP: controle positivo. CN: controle negativo. MP: marcador de peso molecular 1 kb.

As plantas que foram inoculadas com o PMeV resultaram positivas nos três tempos avaliados, com exceção da P02 que foi negativa nos três meses de idade e não conservou o mesmo padrão de amplificação que as outras duas repetições biológicas (P16 e P22). Os resultados do diagnóstico são apresentados na Tabela 5.

Tabela 5: Diagnostico da presença de PMeV em plantas de *C. papaya*.

Planta	2 meses	Tratamentos	3 meses	6 meses	8 meses
03	(-)	<b>Não Inoculadas</b>	(-)	(-)	(-)
08	(-)		(-)	(-)	(-)
33	(-)		(-)	(-)	(-)
02	(-)	<b>Inoculadas</b>	(-)	(+)	(+)
16	(-)		(+)	(+)	(+)
22	(-)		(+)	(+)	(+)

## 4.2. Sequenciamento

A Tabela 6 apresenta os resultados do sequenciamento e o mapeamento das leituras no genoma de *C. papaya*.

Quando o *download* do genoma de *C. papaya* foi feito, foram encontrados 5901 *contigs* e 3208 *supercontigs*. Porém, não foi possível usar a informação contida nos *contigs* para o mapeamento porque o programa *TopHat* gera um erro de índice ao tentar usar os *contigs* como referência.

Um total de 43.9 bilhões de pb foram sequenciados e 435.4 milhões de leituras com uma longitude de 109 pb foram geradas a partir de 24 amostras corridas em uma linha. A soma das leituras entre as réplicas para cada condição oscilou entre os 52 e 58 milhões, porém apenas uma porcentagem entre o 61 e 63% conseguiu ser mapeado no genoma de *C. papaya*. Uma porcentagem bem pequena dentro das leituras não mapeadas pode pertencer ao vírus, mas ainda, aproximadamente, 35% das leituras está por explorar. Os resultados do mapeamento conservaram homogeneidade o qual é um bom indicativo de confiabilidade nos dados obtidos.

Em relação com a qualidade das leituras produzidas no sequenciamento, três parâmetros são mostrados: o Q20, o Q30 e o conteúdo de GC%. O Q20 nos indica a porcentagem das leituras com um 99% ou mais de acurácia (do inglês *accuracy*), ou seja, a probabilidade de que uma vez em 100 vezes uma base seja atribuída de forma errada nos picos do cromatograma durante o sequenciamento (do inglês *base calling*). Por sua parte, o resultado de Q30 representa 99.9% ou mais de acurácia, ou seja, a probabilidade de que uma vez em 1000 vezes uma base seja atribuída de forma errada. Quando um sequenciamento alcança um *phred quality* Q30, praticamente todas as leituras serão perfeitas, com zero erros e ambiguidades (ILLUMINA, 2011). Baseados nestas definições observou-se que, aproximadamente, 97% das leituras geradas neste experimento, em cada condição, têm uma confiabilidade de 99% ou mais, bem como, aproximadamente, 92% das leituras têm uma acurácia de 99.9%. Estes valores nos indicam um alto grau de confiança nas sequências geradas para as análises seguintes.



Tabela 6: Estatística das leituras geradas e seu mapeamento no genoma de *C. papaya*.

	Tempo (meses)	Total Leituras	Total Nucleotídeos	Leituras Mapeadas <sup>1</sup>	%pC <sup>2</sup>	GC% conteúdo	Q20 %	Q30%
Referência	2	52.006.434	5.252.649.834	32.254.898	62,0	44,0	96,9	92,2
Não-Inoculadas	3	55.120.908	5.567.211.708	34.662.905	62,9	44,4	96,8	92,1
	6	53.903.142	5.444.217.342	33.173.355	61,5	44,1	96,9	92,2
	8	54.021.076	5.456.128.676	32.953.232	61,0	44,5	96,8	92,0
Referência	2	58.215.746	5.879.790.346	35.204.030	60,5	44,7	96,8	92,0
Inoculadas	3	54.064.794	5.460.544.194	32.985.193	61,0	44,9	96,8	91,8
	6	54.178.476	5.472.026.076	34.444.110	63,6	43,3	96,8	92,3
	8	53.926.052	5.446.531.252	33.130.776	61,4	44,4	96,9	92,3

1. Número total de leituras mapeadas no genoma de *C. papaya*.

2. Porcentagem das leituras mapeadas no genoma de *C. papaya*.

Por outro lado, tem sido reportado que para dados gerados em plataformas de sequenciamento que usam o sistema de *Illumina*, o viés (do inglês *bias*) no conteúdo de GC, sequências ricas ou sequências pobres no conteúdo destas bases, induzem alinhamentos irregulares ou não alinhamentos das leituras no genoma (CHEN *et al.*, 2013). A porcentagem média no conteúdo de CG obtidas em nosso projeto oscilaram entre 43,3 e 44,9%. Para dados procedentes de uma biblioteca com distribuição normal, como é nosso caso, o esperado é um conteúdo de GC nas leituras, também ajustadas a uma distribuição normal, com médias entre 40-60% (BABRAHAM INSTITUTE, 2014) o que nos indica que nossos dados estão dentro destes limites.

#### 4.3. Correlação entre as amostras

Para analisar a relação entre os níveis de expressão dos genes em todas as amostras nos diferentes intervalos de tempo, duas análises foram feitas: (i) análise de escalonamento multidimensional (do inglês *mutidimensional scaling* - MDS), das distâncias entre cada par de amostras dadas pelo coeficiente de variabilidade biológica (do inglês *biological coefficient of variation* - BCV) Figura 9. (ii) Matriz de distâncias baseadas no método de divergência *Jensen-Shannon* (JS) representadas no *heat map* (apresentação gráfica de dados, onde valores individuais de uma matriz são representados em cores) (Figura 10). Estas análises tiveram o objetivo de encontrar as similaridades ou as diferenças entre os níveis de expressão dos genes de todas as amostras a serem comparadas nas análises de expressão diferencial de genes. Estes tipos de análises são úteis para encontrar as fontes de variação (do inglês *sources of variation*) que possam conduzir a testes estatísticos com um maior número de falsos positivos ou falsos negativos causados por amostras com alto desvio padrão.

De modo geral, as duas representações gráficas mostram as distâncias biológicas que deveriam ser observadas entre diferentes estádios de desenvolvimento. Nas análises, revelou-se uma alta correlação entre a maioria de amostras que pertencem a um mesmo estágio de desenvolvimento. No *heat map* apresentado na Figura 10 foi observado que cinco das amostras nos dois meses se encontram agrupadas e só uma delas é distante. Nas amostras de plantas com seis meses observou-se o

mesmo padrão que aos dois meses, onde cinco amostras são altamente similares e uma amostra se distancia do grupo. A maior correlação entre amostras do mesmo tempo de desenvolvimento foi achada nos oito meses, onde todas as seis plantas estão altamente correlacionadas, e a menor correlação foi observada entre as plantas que se encontram nos três meses, onde duas são muito distantes. De modo interessante, as amostras nos dois meses resultaram ser mais similares com as amostras dos seis meses e as amostras dos três meses mais similares com as amostras dos oito meses.

Apesar de que nos três e oito meses para algumas condições, uma alta correlação foi encontrada nas três repetições do mesmo tratamento, este não é um padrão geral e bem se encontram repetições que se distanciam das outras duas ou se encontra uma baixa correlação entre as três. Por conseguinte, este resultado mostra a necessidade de usar réplicas para a comparação entre os dois tratamentos durante os três estádios de tempo avaliados.

A associação entre uma alta correlação entre amostras que se encontram num mesmo ponto de desenvolvimento também foi reportada por (DE CREMER *et al.*, 2013) que fizeram um estudo transcriptômico baseado no RNA-seq com plantas de *Lactuca sativa* infetadas e não infectadas com o fungo *Botrytis cinerea* nas 12, 24 e 48 horas após da inoculação e observou que as amostras tomadas nas 12 horas são dissimilares e apresentam uma maior correlação que as amostras tomadas nos 24 e 48 horas.

#### **4.4. Análises de expressão diferencial de genes**

Nas análises de expressão diferencial, as plantas não inoculadas e inoculadas foram avaliadas ao longo dos três tempos, cada uma comparada com suas correspondentes três repetições biológicas nos dois meses de idade, que são plantas que se encontravam no estágio de prefloração e que não foram infetadas com o PMeV. Os genes foram considerados como diferencialmente expressos quando seu valor q foi igual ou menor que 0,05.

Na Tabela 7 e Tabela 8 são mostrados os genes que foram expressos

diferencialmente aos três, seis e oito meses de idade nas plantas não inoculadas e nas plantas inoculadas respectivamente. Aos três meses nas plantas inoculadas um número maior de genes, foram expressos diferencialmente, observou-se que aproximadamente 2000 genes mais do que nas plantas não inoculadas no mesmo tempo de desenvolvimento, foram expressos. Nos seis e oito meses de desenvolvimento, não foram observadas grandes diferenças entre os dois tratamentos. Um total de 898 e 991 novos transcritos foram expressos diferencialmente nas plantas não inoculadas e inoculadas respectivamente, nos diferentes tempos.

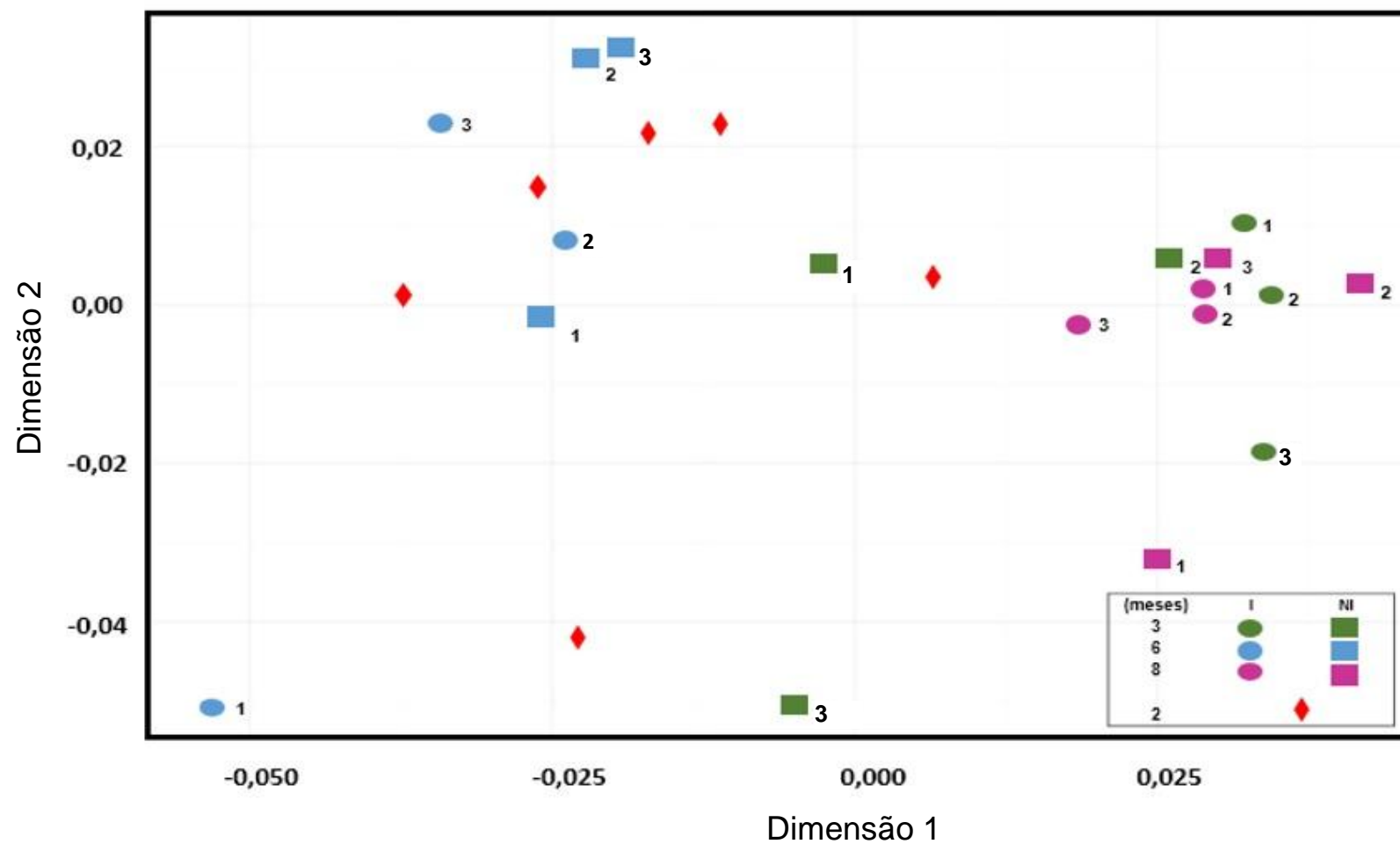


Figura 9: Gráfico de escalonamento multidimensional (MDS) representando as distancias dos coeficientes de variabilidade biológica (BCV) entre todas as amostras. Os círculos representam às plantas inoculadas e os quadrados as plantas não inoculadas. A cor verde representa as plantas com três meses de idade, a cor azul com seis meses de idade e a cor rosa as plantas com oito meses de idade.

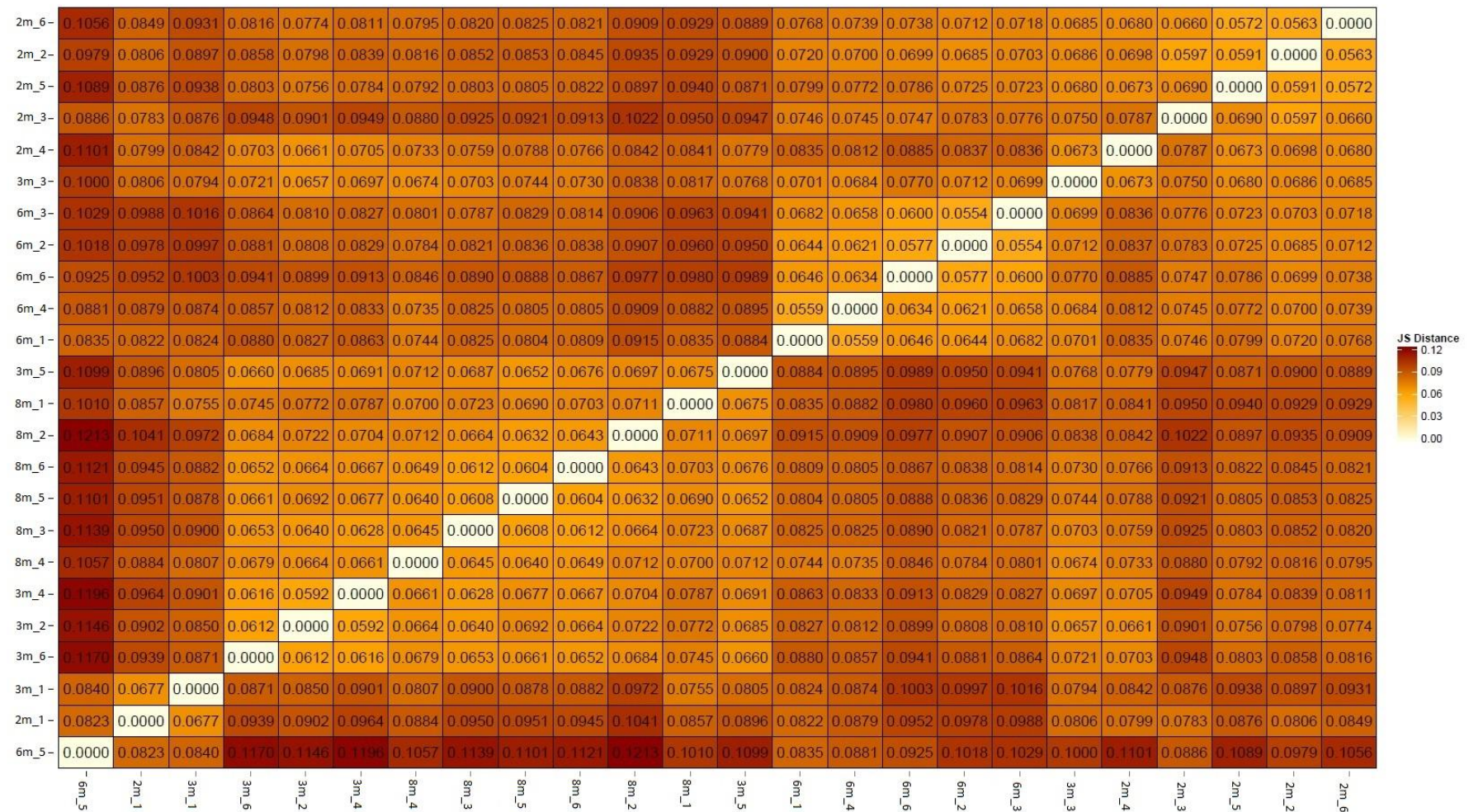


Figura 10: *Heat map* representando as distancias *Jensen-Shannon* (JS) dos níveis de expressão entre todas as amostras sequenciadas.

Observaram-se genes que foram expressos diferencialmente apenas em algum dos tempos e que não foram expressos aos dois meses; estes genes são definidos pelo programa *cuffdiff* como *inf*. Entretanto, outros genes se expressam de modo diferencial somente aos dois meses e não nos outros tempos e são reportados como *-inf*.

Tabela 7: Genes diferencialmente expressos (DEGs) em plantas não Inoculadas com PMeV aos 3,6 e 8 meses de idade quando comparadas com plantas de 2 meses de idade, usando um p-value ajustado a uma FDR=0,05 (q-value).

Idade da planta (Meses)	Genes e transcritos	Induzidos	Reprimidos	Inf	-inf	Total de Genes e transcritos DE	Novos transcritos DE
3	28766	872	739	7	5	1623	104
6	28739	1785	1780	11	9	3585	370
8	28808	2459	2203	28	28	4718	424

Tabela 8: Genes diferencialmente expressos (DEGs) em plantas Inoculadas com PMeV aos 3,6 e 8 meses de idade quando comparadas com plantas de 2 meses de idade, usando um p-value ajustado a uma FDR=0,05 (q-value).

Idade da planta (Meses)	Genes	Induzidos	Reprimidos	Inf	-inf	Total Genes e transcritos DE	Novos transcritos DE
3	28656	1792	1799	9	14	3614	274
6	28883	1566	2097	21	10	3694	382
8	28651	1833	2195	17	13	4058	335

#### 4.5. Gráficos de Análises de expressão diferencial de genes

Para fazer uma análise exploratória da distribuição dos genes e transcritos diferencialmente expressos, dois gráficos são mostrados e descritos a continuação.

Um gráfico de densidade (i) (Figura 11) foi utilizado. Neste gráfico a densidade de genes e transcritos em cada tempo e tratamento é analisada em função dos níveis de expressão dados pela razão *log2 fold change*. Observou-se que nas seis condições avaliadas a maior densidade de genes e transcritos encontram-se nos níveis de expressão compreendidos entre os intervalos (0 e -2,5); (0 e 2,5).

Um diagrama de caixa (ii) foi utilizado tanto para analisar a simetria da distribuição dos dados como para compará-la entre as duas condições e os diferentes tempos. Este tipo de análise usado na estatística descritiva é resumido na

Tabela 9. Este tipo de gráfico proporciona informação relacionada com a mediana, os quartis e os dados atípicos. O quartil inferior contém o 25% dos dados e o quartil superior o 75%. A linha dentro da caixa representa a posição da mediana. O diagrama de caixa representado na Figura 12 nos mostra a distribuição de genes nas duas condições aos três, seis, e oito meses em relação com seus níveis de expressão. Foi observado que nas plantas não inoculadas em todos os tempos as medianas estão localizadas acima do zero, portanto mais do 50% dos dados se encontram com níveis de expressão positivos indicando que a maioria de genes e transcritos são superexpressos. Nas plantas inoculadas em todos os tempos as medianas estão localizadas abaixo do zero, portanto mais do 50% dos dados se encontram com níveis de expressão negativos, indicando que a maioria de genes e transcritos estão sendo reprimidos.

Apesar de o mamão ser uma planta susceptível à infecção pelo PMeV, ou seja, a planta não consegue deter a infecção, desencadeando finalmente os sintomas após o florescimento, os resultados observados apontam a uma resposta de defesa primaria disparada nos três meses de idade, quando um número consideravelmente maior de genes são expressos diferencialmente nas plantas inoculadas do que nas plantas não inoculadas. Estes genes diferencialmente expressos neste estágio deverão ser atentamente analisados em futuros trabalhos.



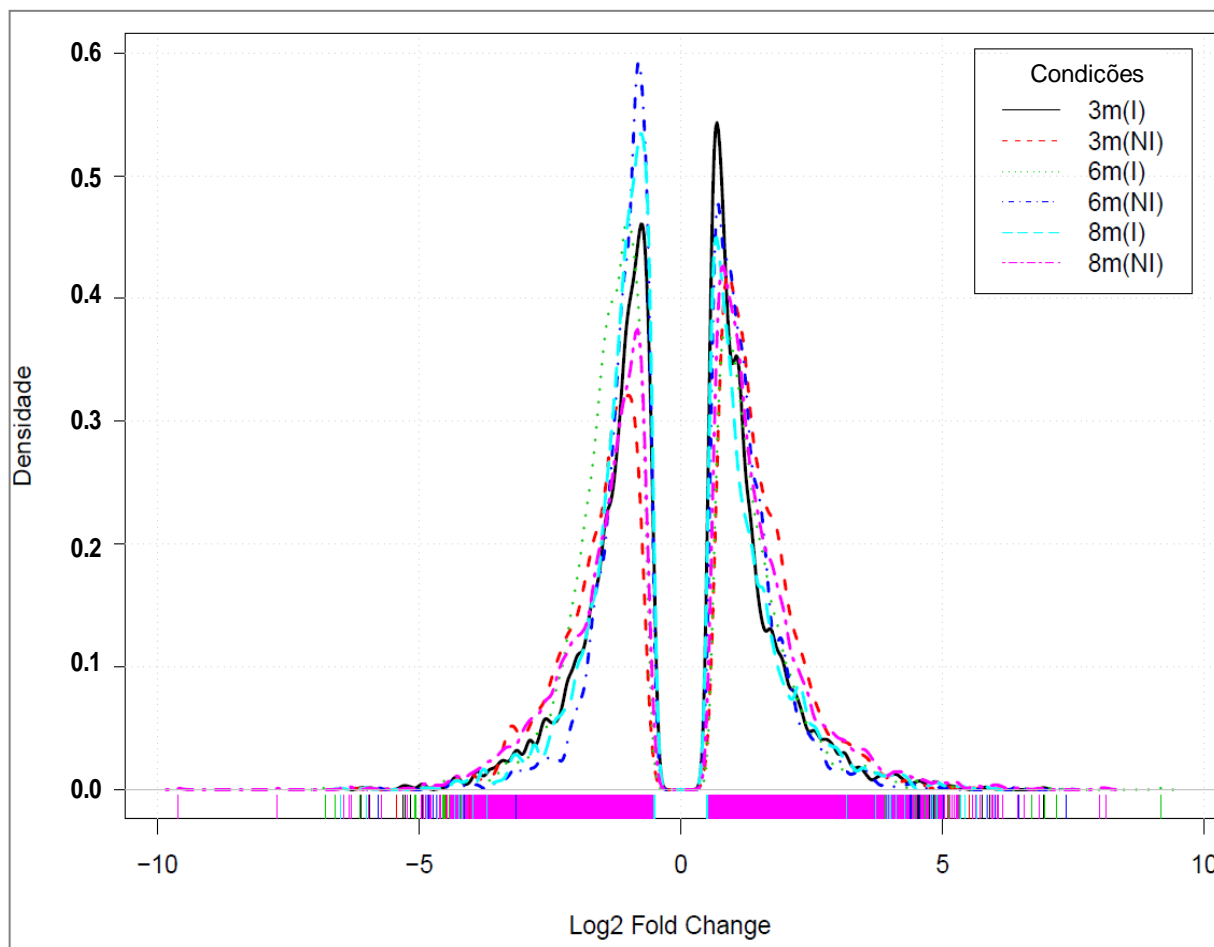


Figura 11: Gráfico de densidade. Densidade de genes em função do seus níveis de expressão dados pela razão Log2 (Fold Change) para todas as condições.

Além desta resposta primária, os resultados mostrados na Figura 12 sugerem um padrão de resposta geral das plantas inoculadas dada durante os três estádios de desenvolvimento avaliados na qual a maioria de genes e transcritos estão sendo reprimidos. Entretanto, nas plantas não inoculadas, a maioria dos genes transcritos encontram-se superexpressos. Este resultado adquire sentido se levarmos em consideração que as plantas inoculadas se encontram submetidas a uma forte condição de estresse, e que em condições de estresse as plantas reduzem a taxa de alguns processos fisiológicos como crescimento ou fotossínteses abaixo da taxa normal. Esta resposta imediata na qual a planta reduz seu desempenho é feita para compensar os efeitos prejudiciais causadas pelo estresse (LAMBERS *et al.*, 2008).

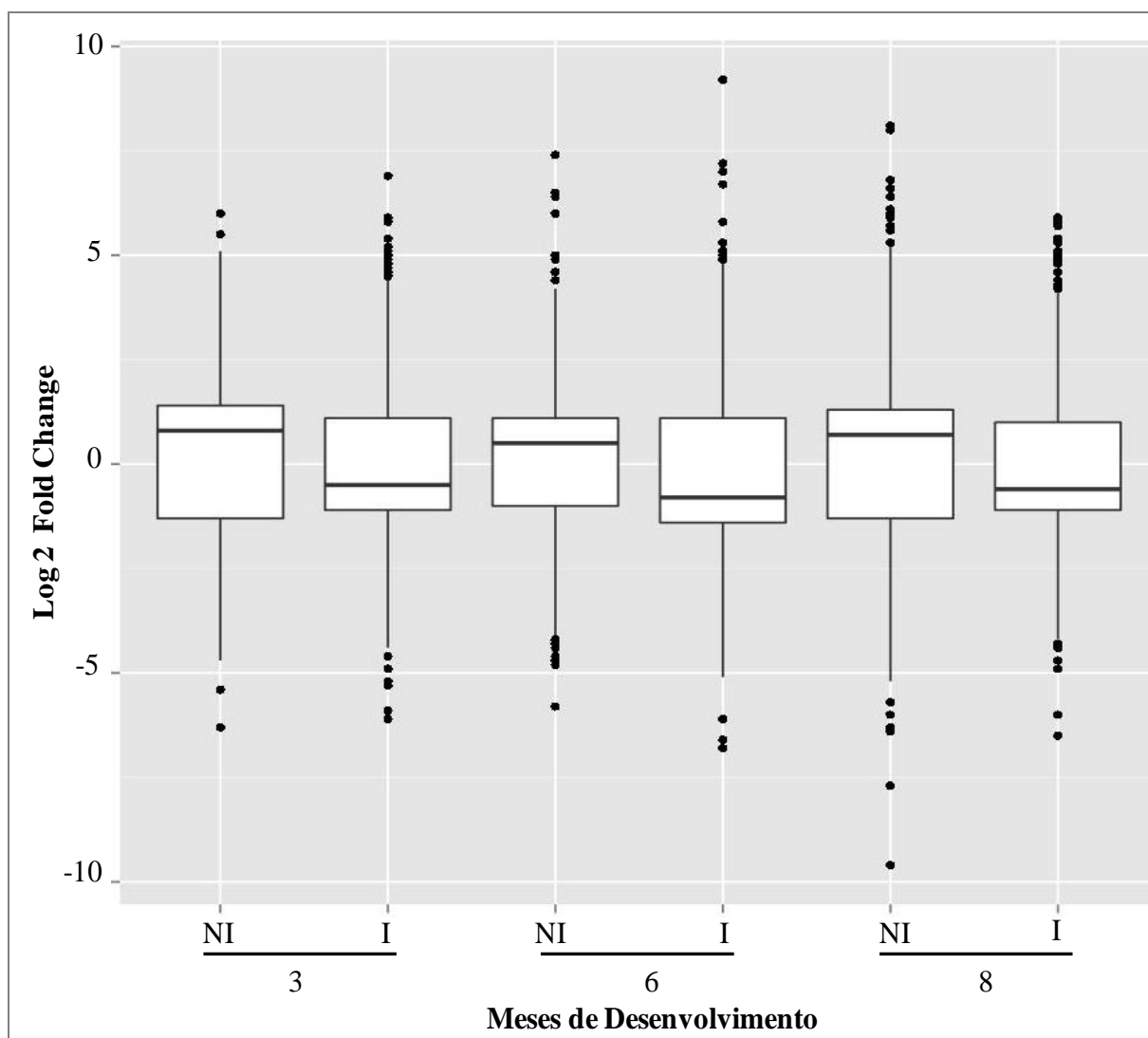


Figura 12: Diagrama de caixas. Distribuição dos genes diferencialmente expressos nas plantas não inoculadas (NI) e inoculadas (I) nos três, seis e oito meses de idade. em relação aos níveis de expressão dados pela razão *Log2 Fold change*.

Tabela 9: Sumario da distribuição estatística dos dados no diagrama de caixa.

Meses	Mediana	Sd	IQR	0%	25%	50%	75%	100%	Dados
3(NI)	0,08	1,8	2,8	-6,3	-1,3	0,8	1,4	6,0	1611
3(I)	-0,02	1,6	2,2	-6,1	-1,1	-0,5	1,1	6,9	3591
6(NI)	0,06	1,4	2,1	-5,8	-1,0	0,5	1,1	7,4	3565
6(I)	-0,22	1,6	2,5	-6,8	-1,4	-0,8	1,1	9,2	3663
8(NI)	0,07	1,8	2,6	-9,6	-1,3	0,7	1,3	8,1	4662
8(I)	-0,06	1,5	2,1	-6,5	-1,1	-0,6	1,0	5,9	4028

A ativação de uma resposta primária e geral na interação *C. papaya*-PMeV tem já sido sugerida em outros trabalhos e é exposta no primeiro capítulo deste documento. Principalmente, estes estudos mostraram que nas plantas infectadas pelo PMeV é observado um acúmulo de cristais de oxalato de cálcio e, o aumento da produção de peróxido de hidrogênio (H<sub>2</sub>O<sub>2</sub>) (RODRIGUES *et al.*, 2009) e, a alteração em miRNAs que modulam genes envolvidos em respostas de defesa ao estresse como a via de espécies reativas de oxigênio (ROS) (ABREU *et al.*, 2014). Estes processos evocam uma resposta de defesa similar a HR mas que contrasta com HR no fato de que esta resposta é própria de interações não compatíveis que levam à resistência do hospedeiro (MOREL e DANGL, 1997; HEATH, 2000; MUR *et al.*, 2008).

Alternativamente a HR, outro tipo de resposta que envolve mecanismos fisiológicos, bioquímicos e moleculares que se assemelham a HR, mas que ocorre nas interações compatíveis, chamada necrose sistêmica, exposta em (MANDADI e SCHOLTHOF, 2013), poderia estar sendo manifestada na interação *C. papaya*-PMeV.

Embora que não se tinha esclarecido se esta resposta é uma resposta própria de interações compatíveis ou é uma resposta HR não controlada ou incompleta que é disparada nos tecidos distais quando a HR que é uma resposta de tipo local, não conseguiu evitar a propagação do vírus, este tipo de resposta já tem sido citada em várias outras interações (SCHOLTHOF, 1999; CHU *et al.*, 2000; XU e ROOSSINCK, 2000; GONZALEZ-JARA *et al.*, 2004; OZEKI *et al.*, 2006; MANDADI e SCHOLTHOF, 2012).

Do mesmo modo outras respostas de estresse que envolve o sistema de proteossoma UPS têm sido sugeridas. ABREU *et al.* (2014) observou a acumulação de miRNAs presentes neste sistema e, (RODRIGUES *et al.*, 2011), mostrou que durante a infecção pelo PMeV proteínas relacionadas ao metabolismo, são reprimidas, enquanto que a calreticulina e proteínas da via UPS relacionadas ao estresse, são induzidas durante a infecção.

Todos estes resultados e observações reforçam a hipótese da ativação de respostas de defesa em mamão. A planta tenta se defender, mas, aparentemente, esta

resposta é truncada após a floração. Possivelmente, algum mecanismo molecular é disparado nesta interação, em que alguma via de sinalização envolvida no desenvolvimento fisiológico da planta coaja e ative uma via envolvida também em respostas de defesa e a planta dispare os sintomas como um mecanismo para inativar o vírus. Por outro lado, pode-se inferir que o vírus se aproveite que a planta nesse ponto investe todos seus recursos para focar-se num processo crucial que envolve seu desenvolvimento, sendo obrigada a reduzir as possíveis respostas que estavam atenuando ou controlando a infecção; com isto os sintomas são disparados e usados pelo vírus como um mecanismo para uma maior disseminação.

Os eventos moleculares envolvidos na floração e pos-floração envolvem complexos mecanismos moleculares que na presença do estresse são seriamente afetados e modificados. Na floração os efeitos causados pelo estresse podem ser atribuídos em parte às alterações no epigenoma que tomam lugar no nucleossomo e envolvem modificações na cromatina. Alterações no epigenoma estão fortemente ligadas com as alterações nos padrões de expressão gênica no desenvolvimento da floração que tem sido observadas no estresse biótico bem como no estresse abiótico (ALVAREZ *et al.*, 2010; CHEN *et al.*, 2010).

Para encontrar um perfil de expressão entre os genes que fossem comuns entre as duas condições e em todos os estádios de desenvolvimento, bem como poder determinar os efeitos da infecção pelo PMeV em mamão diferentes análises foram feitas. (1) Os DEGs foram analisados usando diagramas de Venn (Figura 13). Este tipo de análise, além de mostrar os genes que são exclusivos de uma condição particular, mostra também, os genes são compartilhados. Na figura 15a se resalta que 596 genes foram comunmente expressos nos três tempos avaliados nas plantas não inoculadas e do mesmos modo que na figura 15b se ressaltam os 910 genes são comunente expressos durante todos os estádios de desenvolvimento nas plantas não inoculadas e inoculadas. Na Figura 15c se mostra a relação entre os 596 e 910 genes comunente expressos nas plantas não inoculadas e inoculadas respectivamente, onde 331 genes são expressos exclusivamente nas plantas não inoculadas, 645 exclusivamente nas plantas não inoculadas e, 265 genes se expressam nos dois tratamentos e nos três tempos avaliados.

O diagrama de Venn nos permitiu determinar que 265 genes se expressam

independentemente ao tratamento e aos tempos avaliados. Porém para conhecer como estão sendo expressos estes genes, e saber se os níveis de expressão são alterados ou permanecem iguais nos dois tratamentos e nos três tempos, uma outra análise (2) complementar com a anterior (Figura 16) deveu ser abordada. Esta análise de *clusters* permitiu visualizar o perfil de expressão destes 265 genes.

Na Figura 14a, observou-se que a maioria dos genes possuem um mesmo padrão nos níveis de expressão ao longo de todos os tempos e condições, no entanto 8 *clusters* (Figura 14b) mostraram um padrão diferencial, principalmente nos seis meses de desenvolvimento, onde alguns *clusters* apresentam um padrão diferencial nas duas condições, e, que provavelmente correspondem com genes que se expressam diferencialmente nos seis meses de idade mas que sua diferença é devida a processos fisiológicos que acontecem exclusivamente nesse estágio do desenvolvimento da planta. Entretanto os clusters I, III e, VII contêm genes que mostram um padrão de expressão diferente somente nas plantas inoculadas com seis meses de idade. Nestes três *clusters*, oito genes foram encontrados reprimidos e treze genes superexpressos. O padrão de expressão observado aponta a que a resposta que é crucial no desenvolvimento dos sintomas na planta provavelmente ocorre nos seis meses de desenvolvimento.

Os genes dos *cluster* I, III e, VII foram anotados usando *Gene ontology* (GO). A Figura 15 mostra a classificação por processo biológico dos oito genes que foram reprimidos e dos treze genes superexpressos. Os oito genes reprimidos estão envolvidos em 13 processos biológicos; por outro lado, os 13 genes superexpressos estão envolvidos em 65.

Conforme os resultados da classificação feita por *Gene ontology* em relação aos processos biológicos, podemos observar que genes envolvidos em processos como morte celular, resposta aos diferentes estímulos biótico e abiótico, bem como resposta ao estresse, processos biossintético e metabolismo celular se encontram reprimidos, enquanto que genes envolvidos em processos metabólicos primários, biogêneses, processos celular de organismo único e multicelular (do inglês *single-organism and multicellular process*) que envolve processos de diferenciação celular, ciclo celular, comunicação celular e crescimento celular, bem como processos envolvidos em reprodução, e desenvolvimento floral, se encontram induzidos. Este

resultado sugere que neste ponto do desenvolvimento as plantas mostram uma ativação de processos que envolvem principalmente metabolismo, reprodução, floração e processos celulares e a diminuição de processos relacionados com a resposta de defesa e, resposta ao estresse.

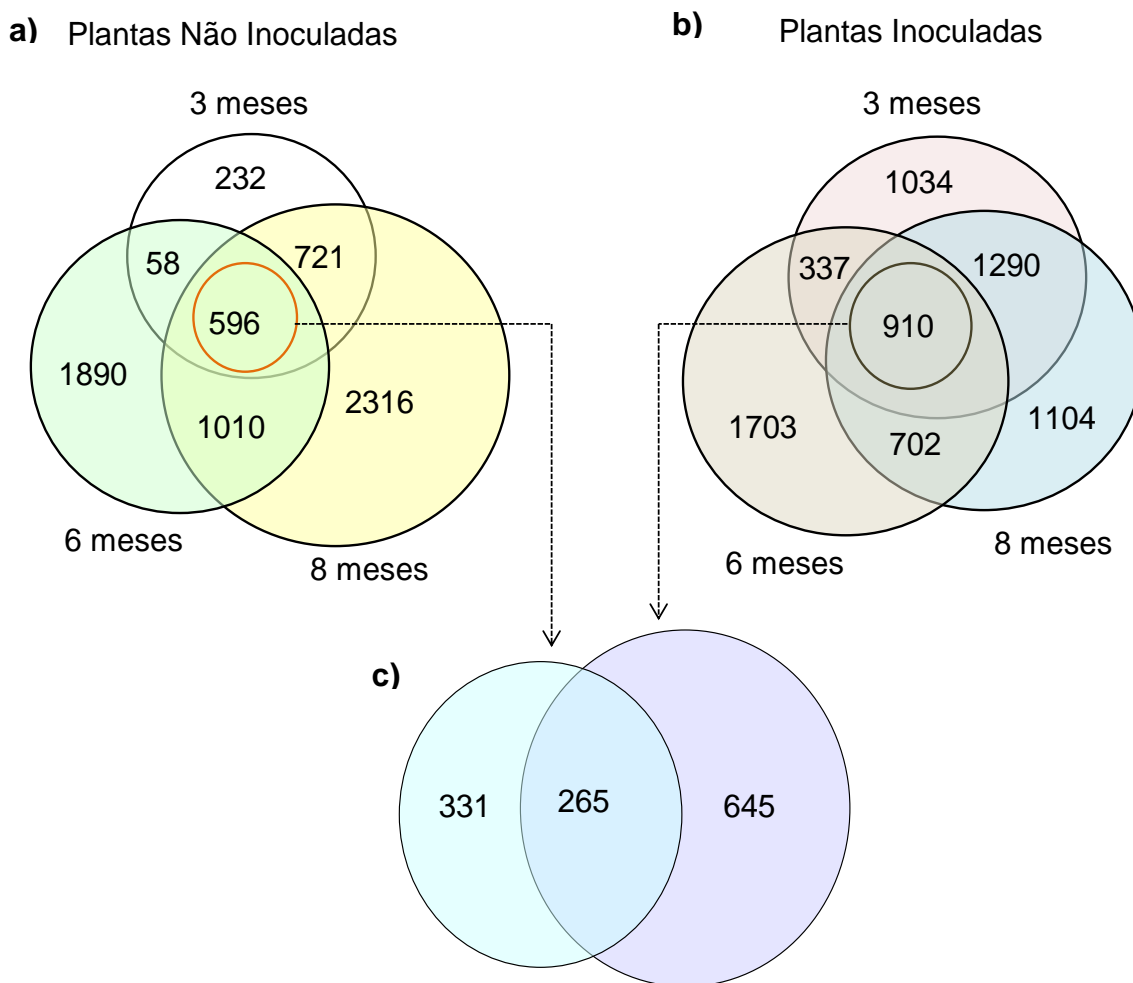
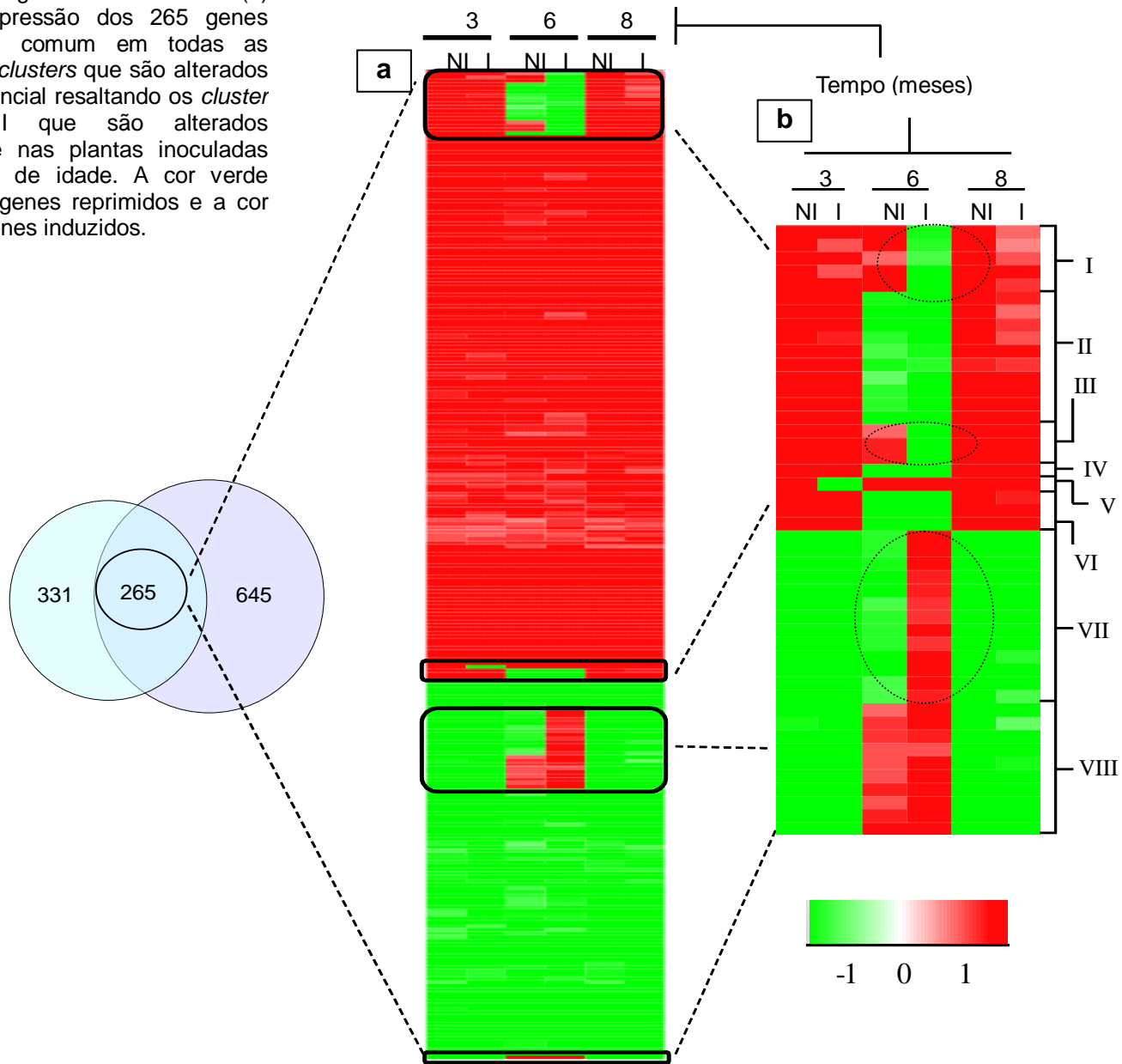


Figura 13: Diagramas de Venn mostrando as relações entre os diferentes conjuntos de genes diferencialmente expressos nas (a) plantas não inoculadas e (b) plantas inoculadas aos 3, 6, e 8 meses de idade quando comparados com plantas de 2 meses de idade. (c) Relação entre os genes que são comumente expressos nas plantas inoculadas e não inoculadas em todos os tempos avaliados, se mostram os genes que se expressam exclusivamente nas plantas inoculadas, não inoculadas e, os genes que são comumente expressos nos dois tratamentos.

Figura 14: Diagrama de cluster. (a) padrão de expressão dos 265 genes expressos em comum em todas as condições. (b) *clusters* que são alterados de modo diferencial ressaltando os *cluster* I, III e VII que são alterados exclusivamente nas plantas inoculadas com 6 meses de idade. A cor verde representa os genes reprimidos e a cor vermelha os genes induzidos.



condções.

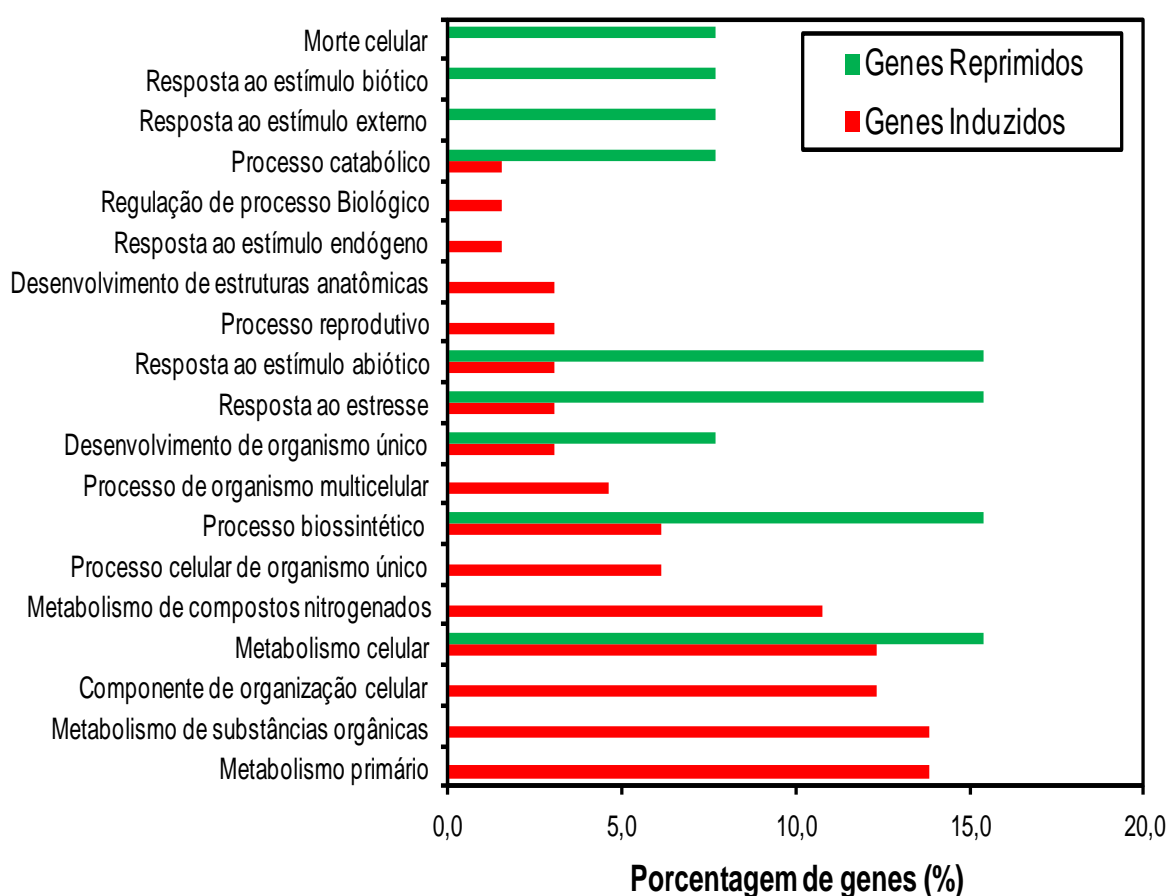


Figura 15: Anotação funcional GO dos genes. Distribuição dos genes presentes nos *clusters* I, III e VII associados com os diferentes processos biológicos. A cor verde representa os genes reprimidos e a cor vermelha os genes induzidos nos seis meses de idade.

Estes resultados mostram que nos seis meses de desenvolvimento as plantas estão reprogramando sua expressão de genes e ativando uma resposta que contrasta com as respostas de defesa observadas nos outros tempos de desenvolvimento, conforme discutido anteriormente, onde foram reportadas a diminuição das funções do metabolismo e aumento de funções do estresse (RODRIGUES *et al.*, 2011).

Nas plantas que foram inoculadas com o vírus, neste ponto, provavelmente são forçadas a reduzir os recursos que estavam comprometidos com o balance entre seu desenvolvimento e a resposta ao estresse, primando a expressão de genes envolvidos no desenvolvimento de processos reprodutivos, metabólicos e biossínteses de constituintes de macromoléculas. Esta reprogramação na expressão de genes favorece a propagação da infecção pelo vírus levando finalmente ao desenvolvimento da doença.

Poucos trabalhos que abordem o padrão geral de expressão de genes durante o



desenvolvimento de mamão tem sido feitos e, ainda menos que abordem sua interação planta patógeno. No entanto (FABI *et al.*, 2012), usando a tecnologia do microarranjo, mostraram respostas celulares durante a maturação, em que genes relacionados com o estresse abiótico são aumentados e genes relacionados ao estresse biótico são diminuídos. (PORTER *et al.*, 2008) fizeram um estudo transcriptômico nas raízes de mamão que revelou genes associados a uma complexa rede de interações que incluem defesa, interações que favorecem a interação planta-microbio, estresse abiótico e desenvolvimento.

A reprogramação na expressão de genes sob condições de estresse têm sido amplamente discutida. Quando um organismo é submetido a um stress, as vias de sinalização que controlam quase todos os processos da fisiologia celular são ativadas. A modificação na expressão de genes junto com as alterações no metabolismo, progressão do ciclo celular, homeostases de proteínas, organização do citoesqueleto, tráfico vesicular, e, modificação de atividades enzimáticas, são as principais componentes na resposta ao stress (WESTFALL, 2004; HOHMANN *et al.*, 2007; GEHART *et al.*, 2010; RICHTER *et al.*, 2010).

A resposta a estresse tem sido amplamente estudada usando como modelo *Saccharomyces cerevisiae* (GASCH *et al.*, 2000; CAUSTON *et al.*, 2001; CHEN *et al.*, 2003). Estes trabalhos mostram como a pesar de que diferentes estresses geram respostas específicas, uma resposta geral e, amplamente conservada é observada. Nesta resposta, os genes de choque térmico funções antioxidantes, bem como metabolismo de carboidratos e geração de energia são ativados, enquanto que genes relacionados com funções do crescimento celular como translação (do inglês *translation*) e biogêneses dos ribossomos são reprimidos. Estas observações mostram como neste modelo os recursos destinados para uma rápida proliferação são redirecionados para uma proteção frente ao estresse.

De aqui podemos inferir que para estes organismos é essencial manter um balance entre a programação na sua expressão de genes relacionados com o crescimento celular ou com a resposta ao estresse. Este balance depende é influenciado diretamente pelos fatores ambientais. As células devem estar adaptadas e ter mecanismos de regulação que permitam sua reprogramação em qualquer momento.

As plantas adaptam seu crescimento e processos de desenvolvimento em resposta às condições ambientais. Em condições de estresse processos fisiológicos encaminhados à redução dos danos celulares são induzidos ao mesmo tempo que as plantas devem alterar seu tempo de desenvolvimento para completar seu ciclo de vida em um tempo oportuno por conseguinte plantas submetidas ao estresse experimentam uma transição para seus estádios reprodutivos mais cedo do que as plantas não submetidas ao estresse (YAISH *et al.*, 2011).

Estes mecanismos moleculares e fisiológicos discutidos; leva-nos a concluir que o florescimento é um estágio complexo que inclui vias e mecanismos moleculares que estão envolvidos paralelamente em diversos processos como na resposta ao estresse. Por conseguinte uns processos podem interferir com os outros de modo positivo ou negativo sempre tentando manter o equilíbrio para a planta.

Nosso trabalho, que é pioneiro no estudo de transcricção na interação planta-patógeno em plantas de ciclo de vida longo como mamão e, que além avalia a evolução da infecção pelo PMeV durante três estádios de desenvolvimento, revela a alteração no padrão de expressão que ocorre em mamão quando é submetida a um estresse de tipo biótico, e como estas respostas são alteradas ao longo da infecção em função do desenvolvimento da planta.

Portanto, corroborando com trabalhos prévios, observou-se que a planta dispara uma resposta de defesa à infecção. Entretanto, quando é avaliada a evolução da resposta ao longo do desenvolvimento da floração até o aparecimento dos sintomas, encontra-se que nos seis meses este padrão de resposta é alterado. Portanto, estes resultados sugerem que, exceto aos seis meses, nos demais tempos avaliados a planta conseguiu manter o balanço entre seus processos de reprodução e crescimento em relação com os processos de resposta ao estresse. Por outro lado, aos seis meses as plantas devem mudar seu programa de expressão gênica e direcionar seus recursos para favorecer os processos reprodutivos e metabólicos que ocorrem especificamente nesse estágio de desenvolvimento e que prevalecem sob a resposta de defesa ao estresse, causando o desenvolvimento dos sintomas.

## 5. CONCLUSÃO

Usando a plataforma de sequenciamento *Illumina* HiSeq 2000 foram gerados 43.9 bilhões de pb em 435.4 milhões de leituras partir de 24 amostras.

Durante os três estádios de desenvolvimento avaliados 9926 e 11366 genes e transcritos expressos diferencialmente nas plantas não inoculadas com o PMeV e plantas inoculadas com PMeV foram identificados respectivamente.

Novos transcritos foram expressos diferencialmente. 898 e 991 novos transcritos foram identificados nas plantas não inoculadas e inoculadas respectivamente.

A análise gráfica dos genes diferencialmente expressos que são comuns aos três estádios de desenvolvimento e nas duas condições, revelou um padrão de expressão geral dos genes envolvidos na interação mamoeiro-PMeV.

Encontraram-se 21 genes com o perfil de expressão alterado nas plantas inoculadas exclusivamente nos seis meses de idade. Destes 21 genes 8 foram encontrados reprimidos e estão envolvidos em processos de respostas de defesa que envolvem morte celular e resposta ao estresse. Os restantes 13 genes são superexpressos e se encontram envolvidos principalmente em processos metabólicos primários, biogêneses, processos que envolvem diferenciação e ciclo celular, comunicação e crescimento celular, bem como processos envolvidos em reprodução, e desenvolvimento floral.

Estes resultados em conjunto com anteriores estudos apontam a que nos seis meses de idade a planta altera seu programa de expressão gênica direcionado para responder a processos próprios do desenvolvimento requeridos nesse estágio fisiológico que primam sob a resposta ao estresse e que levam ao desenvolvimento dos sintomas.

## 6. REFERÊNCIAS

- ABREU, P. et al. Molecular diagnosis of Papaya meleira virus (PMeV) from leaf samples of *Carica papaya* L. using conventional and real-time RT-PCR. **Journal of virological methods**, v. 180, n. 1, p. 11-17, 2012.
- ABREU, P. M. V. et al. *Carica papaya* MicroRNAs Are Responsive to Papaya meleira virus Infection. **PLoS ONE**, v. 9, n. 7, p. e103401, 2014.
- ALABI, O. J. et al. High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease. **Molecular Plant Pathology**, v. 13, n. 9, p. 1060-1076, 2012.
- ALVAREZ, M. E.; NOTA, F.; CAMBIAGNO, D. A. Epigenetic control of plant immunity. **Molecular Plant Pathology**, v. 11, n. 4, p. 563-576, 2010.
- ARYAL, R. et al. Asymmetric purine-pyrimidine distribution in cellular small RNA population of papaya. **Bmc Genomics**, v. 13, Dec 5 2012.
- BABRAHAM INSTITUTE. Babraham Bioinformatics. **Projects: FASTQ**, 2014. Disponível em: < <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> >. Acesso em: 20 Nov. 2014.
- BARBA, M.; CZOSNEK, H.; HADIDI, A. Historical perspective, development and applications of next-generation sequencing in plant virology. **Viruses**, v. 6, n. 1, p. 106-36, Jan 2014.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289-300, 1995.
- BUSS, D. S. et al. Oxidative Stress Defence Response of *Carica papaya* Challenged by Nitric Oxide, Papaya meleira virus and *Saccharomyces cerevisiae*. **The Open Nitric Oxide Journal**, v. 3, p. 55-64, 2011.
- CARR, J. P.; LEWSEY, M. G.; PALUKAITIS, P. Signaling in induced resistance. **Adv Virus Res**, v. 76, p. 57-121, 2010.
- CAUSTON, H. C. et al. Remodeling of yeast genome expression in response to environmental changes. **Mol Biol Cell**, v. 12, n. 2, p. 323-37, Feb 2001.
- CORREA, F. J. F. et al. **Estudo preliminar sobre a exsudação de látex do mamoeiro**. SIMPÓSIO BRASILEIRO DA CULTURA DO MAMOEIRO, 2. UNESP: Anais Jaboticabal 409-428 p. 1988.
- CULVER, J. N.; PADMANABHAN, M. S. Virus-Induced Disease: Altering Host Physiology One Interaction at a Time. **Annual Review of Phytopathology**, v. 45, n. 1, p. 221-243, 2007.

CHEN, D. et al. Global transcriptional responses of fission yeast to environmental stress. **Mol Biol Cell**, v. 14, n. 1, p. 214-29, Jan 2003.

CHEN, M.; LV, S.; MENG, Y. Epigenetic performers in plants. **Development, Growth & Differentiation**, v. 52, n. 6, p. 555-566, 2010.

CHEN, Y.-C. et al. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. **PLoS ONE**, v. 8, n. 4, p. e62856, 2013.

CHU, M. et al. Genetic dissection of tomato bushy stunt virus p19-protein-mediated host-dependent symptom induction and systemic invasion. **Virology**, v. 266, n. 1, p. 79-87, Jan 5 2000.

DALTRO, C. B. et al. Genetic diversity studies of Papaya meleira virus. **Tropical Plant Pathology**, v. 39, n. 1, p. 104-108, 2014.

DE CREMER, K. et al. RNAseq-based transcriptome analysis of *Lactuca sativa* infected by the fungal necrotroph *Botrytis cinerea*. **Plant, Cell & Environment**, p. n/a-n/a, 2013.

DRUGEON, G.; JUPIN, I. Stability in vitro of the 69K movement protein of Turnip yellow mosaic virus is regulated by the ubiquitin-mediated proteasome pathway. **Journal of General Virology**, v. 83, p. 3187–3197, 2002.

FABI, J. P. et al. Analysis of ripening-related gene expression in papaya using an Arabidopsis-based microarray. **BMC Plant Biology**, v. 12, n. 1, p. 242, 2012.

FAO. FAOSTAT Statistics Division. 2013. Disponível em: < <http://faostat.fao.org/site/291/default.aspx> >. Acesso em: 8 Nov. 2014.

FAO. FAOSTAT Statistics Division. 2014. Disponível em: < <http://faostat3.fao.org/faostat-gateway/go/to/home/E> >. Acesso em: 8 Nov. 2014.

GAN, X. et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. **Nature**, v. 477, n. 7365, p. 419-423, 2011.

GASCH, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. **Mol Biol Cell**, v. 11, n. 12, p. 4241-57, Dec 2000.

GEHART, H. et al. MAPK signalling in cellular metabolism: stress or wellness? **EMBO Rep**, v. 11, n. 11, p. 834-40, Nov 2010.

GONZALEZ-JARA, P. et al. Host-dependent differences during synergistic infection by Potyviruses with potato virus X. **Molecular Plant Pathology**, v. 5, n. 1, p. 29-35, 2004.

GRAVELEY, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. **Nature**, v. 471, n. 7339, p. 473-479, 2010.

GUSBERTI, M.; GESSLER, C.; BROGGINI, G. A. L. RNA-Seq Analysis Reveals

Candidate Genes for Ontogenic Resistance in Malus-Venturia Pathosystem. **PLoS ONE**, v. 8, n. 11, p. e78457, 2013.

HEATH, M. Hypersensitive response-related death. **Plant Molecular Biology**, v. 44, n. 3, p. 321-334, 2000/10/01 2000.

HERSHKO, A.; CIECHANOVER, A. The Ubiquitin System. **Annual Review of Biochemistry**, v. 67, n. 1, p. 425-479, 1998.

HOHMANN, S.; KRANTZ, M.; NORDLANDER, B. Yeast osmoregulation. **Methods Enzymol**, v. 428, p. 29-45, 2007.

HOWARD, B. E. et al. High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. **PLoS One**, v. 8, n. 10, p. e74183, 2013.

HUA, Z.; VIERSTRA, R. D. The Cullin-RING Ubiquitin-Protein Ligases. **Annual Review of Plant Biology**, v. 62, n. 1, p. 299-334, 2011.

ILLUMINA. **Illumina Sequencing Technology. Highest data accuracy, simple workflow, and a broad range of applications.** Illumina, Inc. USA. 2010

ILLUMINA. **Quality scores for Next-Generation Sequencing: Assessing sequencing accuracy using Phred quality scoring.** Illumina, Inc. 2011

JAIN, P. K. et al. Generation Sequencing Principle and Applications to Crops In: BENKEBLIA, N. (Ed.). **Omics Technologies and Crop Improvement**. New York: CRC Press Taylor & Francis Group, 2014.

JIN, H. L.; LI, S. T.; VILLEGAS, A. Down-regulation of the 26S proteasome subunit RPN9 inhibits viral systemic transport and alters plant vascular development. **Plant Physiology**, v. 142, n. 2, p. 651-661, Oct 2006.

KIM, B. et al. Veinal Necrosis Induced by Turnip mosaic virus Infection in Arabidopsis is a Form of Defense Response Accompanying HR-Like Cell Death. **Molecular Plant-Microbe Interactions**, v. 21, n. 2, p. 260-268, 2008.

KIM, K. H. et al. RNA-Seq Analysis of a Soybean Near-Isogenic Line Carrying Bacterial Leaf Pustule-Resistant and -Susceptible Alleles. **DNA Research**, v. 18, n. 6, p. 483-497, 2011.

KITAJIMA, E. W. et al. Association of isometric viruslike particles, restricted to laticifers, with "meleira" ("Sticky disease") of papaya (Carica papaya). **Fitopatologia Brasileira**, v. 18, p. 118-122, 1993.

KOMATSU, K. et al. Viral-Induced Systemic Necrosis in Plants Involves Both Programmed Cell Death and the Inhibition of Viral Multiplication, Which Are Regulated by Independent Pathways. **Molecular Plant-Microbe Interactions**, v. 23, n. 3, p. 283-293, 2010.

LAMBERS, H.; CHAPIN, F. S.; PONS, T. L. Introduction-History, Assumptions, and Approaches. In: LAMBERS, H., CHAPIN III, STUART F., PONS, THIJS L (Ed.). **Plant Physiological Ecology**. 2: Springer, 2008. p.1-9.

LI, Z. et al. Cdc34p Ubiquitin-Conjugating Enzyme Is a Component of the Tombusvirus Replicase Complex and Ubiquitinates p33 Replication Protein. **Journal of Virology**, v. 82, n. 14, p. 6911-6926, 2008.

LIN, F. et al. Molecular response to the pathogen *Phytophthora sojae* among ten soybean near isogenic lines revealed by comparative transcriptomics. **BMC Genomics**, v. 15, n. 1, p. 18, 2014.

LISTER, R. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. **Nature**, v. 471, n. 7336, p. 68-73, 2011.

LIU, L. et al. Comparison of next-generation sequencing systems. **J Biomed Biotechnol**, v. 2012, p. 251364, 2012.

LIU, P.-P. et al. Interconnection between methyl salicylate and lipid-based longdistance signaling during the development of systemic acquired resistance in *Arabidopsis* and tobacco. **Plant Physiology**, v. 155, p. 1762–1768, 2011.

LORRAIN, S. Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? **Trends in Plant Science**, v. 8, n. 6, p. 263-271, 2003.

LU, J. et al. Transcriptome Analysis of *Nicotiana tabacum* Infected by Cucumber mosaic virus during Systemic Symptom Development. **PLoS ONE**, v. 7, n. 8, p. e43447, 2012.

MACIEL-ZAMBOLIM, E. et al. Purification and some properties of Papaya meleira virus, a novel virus infecting papayas in Brazil. **Plant Pathology**, v. 52, n. 3, p. 389-394, 2003.

MANDADI, K. K.; SCHOLTHOF, K. B. G. Characterization of a Viral Synergism in the Monocot *Brachypodium distachyon* Reveals Distinctly Altered Host Molecular Processes Associated with Disease. **Plant Physiology**, v. 160, n. 3, p. 1432-1452, 2012.

MANDADI, K. K.; SCHOLTHOF, K. B. G. Plant immune responses against viruses: how does a virus cause disease? **Plant Cell**, v. 25, p. 1489–1505, 2013.

MARGUERAT, S.; WILHELM, BRIAN T.; BÄHLER, J. Next-generation sequencing: applications beyond genomes. **Biochemical Society Transactions**, v. 36, n. 5, p. 1091, 2008.

MARTINS, D. D. S. Situação atual da produção integrada de mamão no Brasil. In: MARTINS, D. D. S. (Ed.). **Papaya Brasil: Qualidade do mamão para o mercado interno**. INCAPER, Vitória, 2003. p.97-127.

MARTINS, D. D. S. et al. Interaction between Papaya meleira virus (PMeV) infection

of papaya plants and Mediterranean fruit fly infestation of fruits. **Crop Protection**, v. 36, p. 7-10, 2012.

MARTINS, D. D. S.; VENTURA, J. A.; TATAGIBA, J. D. S. Produção integrada de mamão no Espírito Santo. In: (Ed.). **Produção integrada No Brasil**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2008.

MING, R. et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). **Nature**, v. 452, n. 7190, p. 991-996, 2008.

MIZUNO, H. et al. Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.). **BMC Genomics**, v. 11, n. 1, p. 683, 2010.

MOEDER, W.; YOSHIOKA, K. Lesion mimic mutants: A classical, yet still fundamental approach to study programmed cell death. **Plant Signaling & Behavior**, v. 3, n. 10, p. 764-767, 2008.

MOREL, J. B.; DANGL, J. L. The hypersensitive response and the induction of cell death in plants. **Cell Death Differ**, v. 4, n. 8, p. 671-83, Dec 1997.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature Methods**, v. 5, n. 7, p. 621-628, 2008.

MSKCC. Memorial Sloan-Kettering Cancer Center. Computational Transcriptomics., 2014. Disponível em: < <http://www.raetschlab.org/members/research/transcriptomics> >.

MUR, L. A. J. et al. The hypersensitive response; the centenary is upon us but how much do we know? **Journal of Experimental Botany**, v. 59, n. 3, p. 501-520, 2008.

NAKAGAWA, J.; TAKAYAMA, Y.; SUZUKAMA, Y. Exudação de látex pelo mamoeiro. Estudo de ocorrência em Teixeira de Freitas, BA. **Congresso Brasileiro de Fruticultura**, v. 9, p. 555-559, 1987.

OLIVEROS, J. C. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007. Disponível em: < <http://bioinfogp.cnb.csic.es/tools/venny/index.html> >. Acesso em: 15 Ago. 2014.

OZEKI, J. et al. A single amino acid in the RNA-dependent RNA polymerase of *Plantago asiatica* mosaic virus contributes to systemic necrosis. **Arch Virol**, v. 151, n. 10, p. 2067-75, Oct 2006.

OZSOLAK, F.; MILOS, P. M. RNA sequencing: advances, challenges and opportunities. **Nat Rev Genet**, v. 12, n. 2, p. 87-98, Feb 2011.

PACHECO, R. et al. Comparative Analysis of Transcriptomic and Hormonal Responses to Compatible and Incompatible Plant-Virus Interactions that Lead to Cell Death. **Molecular Plant-Microbe Interactions**, v. 25, n. 5, p. 709-723, 2012.



PALLAS, V.; GARCIA, J. A. How do plant viruses induce disease? Interactions and interference with host components. **Journal of General Virology**, v. 92, n. 12, p. 2691-2705, 2011.

PERAZZOLLI, M. et al. Downy mildew resistance induced by *Trichoderma harzianum* T39 in susceptible grapevines partially mimics transcriptional changes of resistant genotypes. **BMC Genomics**, v. 13, n. 1, p. 660, 2012.

PORTER, B. W. et al. Differentially expressed and new non-protein-coding genes from a *Carica papaya* root transcriptome survey. **Plant Science**, v. 174, n. 1, p. 38-50, 2008.

PROUDFOOT, N. J.; FURGER, A.; DYE, M. J. Integrating mRNA processing with transcription. **Cell**, v. 108, n. 4, p. 501-12, Feb 22 2002.

RADFORD, A. D. et al. Application of next-generation sequencing technologies in virology. **Journal of General Virology**, v. 93, n. Pt\_9, p. 1853-1868, 2012.

REICHEL, C.; BEACHY, R. N. Degradation of Tobacco Mosaic Virus Movement Protein by the 26S Proteasome. **Journal of Virology**, v. 74, n. 7, p. 3330-3337, 2000.

RICHTER, K.; HASLBECK, M.; BUCHNER, J. The Heat Shock Response: Life on the Verge of Death. **Molecular Cell**, v. 40, n. 2, p. 253-266, 2010.

RODRIGUES, C. H.; VENTURA, J. A.; MAFFIA, L. A. Distribuição e transmissão da meleira em pomares de mamão no Espírito Santo. **Fitopatologia Brasileira** v. 14, p. 118, 1989.

RODRIGUES, S. P. et al. Effects of the Papaya meleira virus on papaya latex structure and composition. **Plant cell reports**, v. 28, n. 5, p. 861-871, 2009.

RODRIGUES, S. P. et al. Simplified molecular method for the diagnosis of Papaya meleira virus in papaya latex and tissues. **Summa Phytopathologica**, v. 31, p. 281-283, 2005.

RODRIGUES, S. P. et al. Proteomic analysis of papaya (*Carica papaya* L.) displaying typical sticky disease symptoms. **Proteomics**, v. 11, n. 13, p. 2592-2602, 2011.

RODRIGUES, S. P. et al. Label-free quantitative proteomics reveals differentially regulated proteins in the latex of sticky diseased *Carica papaya* L. plants. **Journal of Proteomics**, v. 75, n. 11, p. 3191-3198, Jun 18 2012.

SCHOLTHOF, K.-B. G. A Synergism Induced by Satellite Panicum Mosaic Virus. **Molecular Plant-Microbe Interactions**, v. 12, n. 2, p. 163-166, 1999/02/01 1999.

SHENDURE, J. The beginning of the end for microarrays? **Nature Methods**, v. 5, n. 7, p. 585-587, 2008.

SOCQUET-JUGLARD, D. et al. Comparative RNA-Seq Analysis of Early-Infected

Peach Leaves by the Invasive Phytopathogen *Xanthomonas arboricola* pv. *pruni*. **PLoS ONE**, v. 8, n. 1, p. e54196, 2013.

SPOEL, S. H.; DONG, X. How do plants achieve immunity? Defence without specialized immune cells. **Nature Reviews Immunology**, v. 12, n. 2, p. 89-100, 2012.

TRAPNELL, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. **Nat Biotechnol**, v. 31, n. 1, p. 46-53, Jan 2013.

TRAPNELL, C.; PACHTER, L.; SALZBERG, S. L. TopHat: discovering splice junctions with RNA-Seq. **Bioinformatics**, v. 25, n. 9, p. 1105-1111, 2009.

TRAPNELL, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 7, n. 3, p. 562-578, 2012.

TRAPNELL, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, v. 28, n. 5, p. 511-515, 2010.

TREMBLAY, A. et al. Gene Expression in Leaves of Susceptible Glycine max during Infection with *Phakopsora pachyrhizi* Using Next Generation Sequencing. **Sequencing**, v. 2011, p. 1-14, 2011.

TRIPATHI, S. et al. Advances in Papaya Genomics In: BENKEBLIA, N. (Ed.). **Omics Technologies and Crop Improvement**. New York: CRC Press Taylor & Francis Group, 2014.

TWINE, N. A. et al. Whole Transcriptome Sequencing Reveals Gene Expression and Splicing Differences in Brain Regions Affected by Alzheimer's Disease. **PLoS ONE**, v. 6, n. 1, p. e16266, 2011.

VENTURA, J.; COSTA, H. Manejo integrado de doenças de fruteiras tropicais: abacaxi, banana e mamão. In: INDEPENDENTE, P. (Ed.). **Manejo Integrado - Produção Integrada - Fruteiras Tropicais - Doenças e Pragas**. 1. Viçosa, MG: UFV, 2003. p.587.

VENTURA, J. A. et al. Meleira do mamoeiro: etiologia, sintomas e epidemiologia. In: MARTINS, D. D. S. (Ed.). **Papaya Brasil: Qualidade do mamão para o mercado interno**. . INCAPER, Vitória, 2003. p.267-276.

VENTURA, J. A.; COSTA, H.; TATAGIBA, J. S. Sintomatologia da meleira do mamoeiro e sua importância para o "roguing". **Fitopatologia Brasileira**, v. 26, p. 536-536, 2001.

VENTURA, J. A.; COSTA, H.; TATAGIBA, J. S. Papaya diseases and integrated control. In: NAQVI, S. A. H. M. (Ed.). **Diseases of Fruits and Vegetables**: Springer, v.II, 2004. p.201-268.

VLOT, A. C.; KLESSIG, D. F.; PARK, S.-W. Systemic acquired resistance: the elusive signal(s). **Current Opinion in Plant Biology**, v. 11, n. 4, p. 436-442, 2008.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57-63, 2009.

WARD, J. A.; PONNALA, L.; WEBER, C. A. Strategies for transcriptome analysis in nonmodel plants. **American Journal of Botany**, v. 99, n. 2, p. 267-276, 2012.

WESTFALL, P. J. When the Stress of Your Environment Makes You Go HOG Wild. **Science**, v. 306, n. 5701, p. 1511-1512, 2004.

XU, L. et al. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. **Journal of Experimental Botany**, v. 62, n. 15, p. 5607-5621, 2011.

XU, P.; ROOSSINCK, M. J. Cucumber mosaic virus D satellite RNA-induced programmed cell death in tomato. **Plant Cell** v. 12, p. 1079–1092, 2000.

XU, P. et al. Genetic Loci Controlling Lethal Cell Death in Tomato Caused by Viral Satellite RNA Infection. **Molecular Plant-Microbe Interactions**, v. 25, n. 8, p. 1034-1044, 2012.

YAISH, M. W.; COLASANTI, J.; ROTHSTEIN, S. J. The role of epigenetic processes in controlling flowering time in plants exposed to stress. **Journal of Experimental Botany**, v. 62, n. 11, p. 3727-3735, 2011.

ZHANG, Z. et al. BSCTV C2 Attenuates the Degradation of SAMDC1 to Suppress DNA Methylation-Mediated Gene Silencing in Arabidopsis. **The Plant Cell**, v. 23, n. 1, p. 273-288, 2011.

## APÊNDICES

### Apêndice A. Controle de qualidade da livreria de cDNA

Amostra	Conc. (ng/ul)	Tamanho do fragmento
P03C1	44.71	274
P08C1	58.51	275
P33C1	56.89	277
P02C1	68.44	278
P16C1	67.53	282
P22C1	61.96	279
P03C3	43.97	271
P08C3	68.72	278
P33C3	75.23	288
P02C3	68.96	292
P16C3	44.46	282
P22C3	62.49	280
P03C7	47.24	281
P08C7	66.63	276
P33C7	76.27	276
P02C7	61.61	273
P16C7	39.33	274
P22C7	59.47	277
P03C11	44.77	276
P08C11	62.05	274
P33C11	64.13	274
P02C11	59.84	266
P16C11	65.37	278
P22C11	71.48	263

## Apêndice B. Fluxograma bioinformático usado para o processamento de dados.

